Explorations in Mathematics - Spring 2017

Dr. Jason Gaddis

Wake Forest University

Welcome!

My name is Dr. Jason Gaddis.

You can often find me in my office, Manchester 338. My office hours are M 10-11, T 3-4, W 10-11, F 11-12, or by appointment.

If you want to contact me, your best bet is email, gaddisjd@wfu.edu.

A little bit about me

I grew up a few hours north of here in Harrisonburg, VA.

A little bit about me

I went to undergrad at Indiana University where I majored in Journalism and Mathematics.

A little bit about me

After college I taught high school in Baltimore City/County for four years.

A little bit about me

I went to the University of Wisconsin - Milwaukee for my PhD. My advisor was Allen Bell.

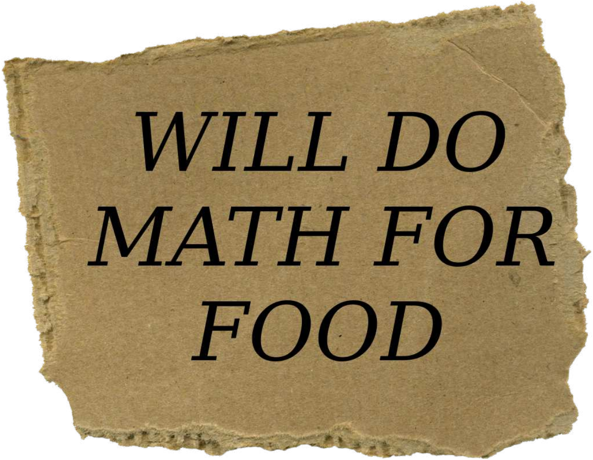
A little bit about me

After graduation, I took a one-year visiting position at the University of California, San Diego.

A little bit about me

Now I am a Teacher-Scholar Postdoctoral Fellow at Wake Forest in my third (and final) year.

A little bit about me



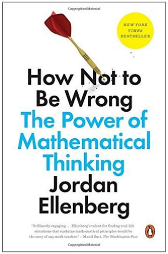
I'm currently "on the market", applying for permanent positions.

This Class

Our class meets MTWF 1:00pm - 1:50pm in Manchester 125. If you didn't know that already you probably wouldn't be here.

Classes will be a mix of discussions, lectures, group activities, and exams. Attendance is of the utmost importance.

Text



Our primary text is *How Not To Be Wrong* by Jordan Ellenberg.

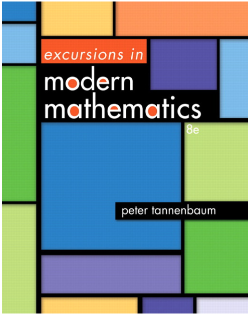
Ellenberg is a professor of mathematics at the University of Wisconsin - Madison.

Text

An optional second text is *Excursions in Modern Mathematics (8e)* by Peter Christmas Tree.

Text

An optional second text is *Excursions in Modern Mathematics (8e)* by Peter ~~Christmas Tree~~ Tannenbaum.



Additional readings will be assigned through the library's course reserve system.

The TA

This is Beth. She is a master's student in the math department and will be the TA for this class.

She will hold study sessions every week on Monday and Thursday night, 7pm-9pm, starting on January

You can also schedule individual appointments with the Math Center.

Grading

At the end of the introduction, Ellenberg writes "..there will be no homework, and there will be no test." This may be true of Ellenberg's book but in this class there will be homework and there will be tests.

Your final grade will be computed using a weighted average on

- Classwork and Participation (10%)
- Reflections (10%)
- Homework (10%)
- Projects (30%)
- Exams (40%) (1/27, 2/17, 3/17, 4/7, 5/3)

Disabilities

If you have a disability which may require an accommodation for taking this course, please contact the Learning Assistance Center (758-5929), then contact me within the first 2 weeks of the semester.

Academic Integrity

Academic integrity is something I take seriously.

In this class I encourage you to talk to other students and myself about assignments. When completing assignments for a grade, however, you must write up/work out the assignment on your own.

During exams, it is expected that you are writing up your own work and do not have any outside help.

Technology

I ask that you silence your cell phone and put all devices away unless specifically instructed otherwise.

What is Math?

This class is about thinking like a mathematician and challenging your intuition.

This class is about exploring mathematical concepts you might not realize even existed.

This class is about answering hard questions with math, or sometimes just figuring out what the question is in the first place.

There are five basic units to this course, each roughly 3 weeks long.

Linearity

Question: Is the obesity epidemic in the US worsening?

Answer: Yes

Question: At some point will everyone in the US be overweight?

Answer: Certainly not.

Inference

Question: Should I invest for retirment?

Answer: Almost certainly.

Question: Should I invest in a mutual fund that looks 'hot'?

Answer: Not necessarily.

Expectation

Question: Will I win the lottery this week?

Answer: No (probably)

Question: Should I play the lottery?

Answer: Maybe.

Regression

Question: Does smoking cause lung cancer?

Answer: Yes!

Question: Does lung cancer cause smoking?

Answer: No.

Existence

Question: Who won the state of Florida in the 2000 presidential election?

Answer: George W. Bush

Question: Who *should have* won the state of Florida in the 2000 presidential election?

Answer: ?

Sequences and Series

Repeated decimals as fractions

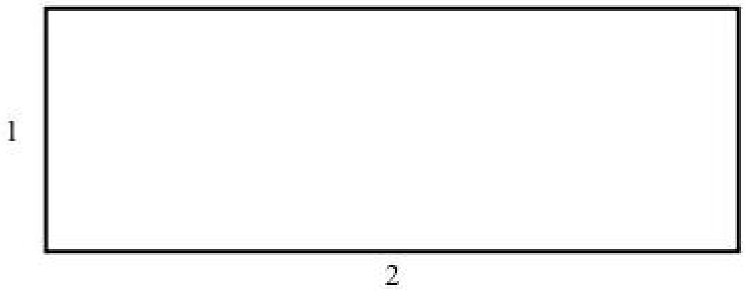
As you will see in the reading, the expression

$$0.9999\ldots$$

is equal to 1.

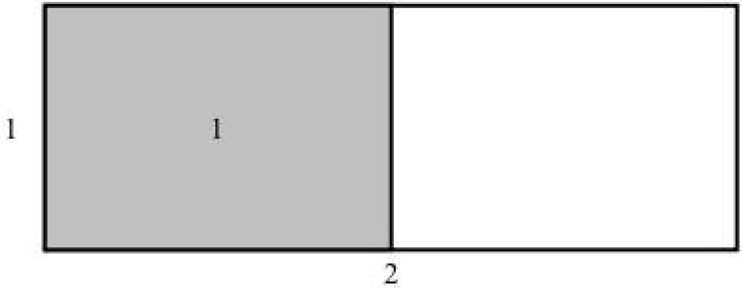
Today we'll see another way to think about this problem and a way to write repeated decimal expressions as fractions.

A geometry problem



By basic geometry, the area of the above rectangle is 2.

A geometry problem



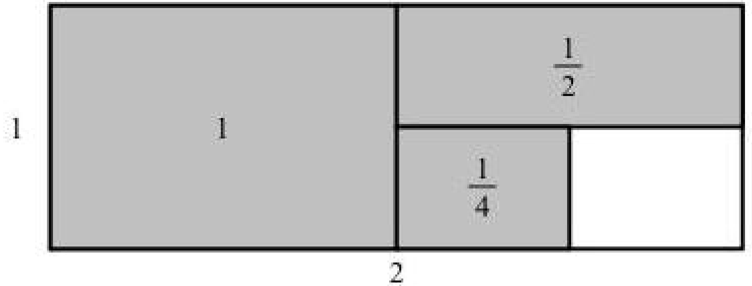
Splitting the rectangle gives two squares, each of area 1.

A geometry problem

After splitting the white rectangle in half, the total shaded area is

$$1+\frac{1}{2}=\frac{3}{2}.$$

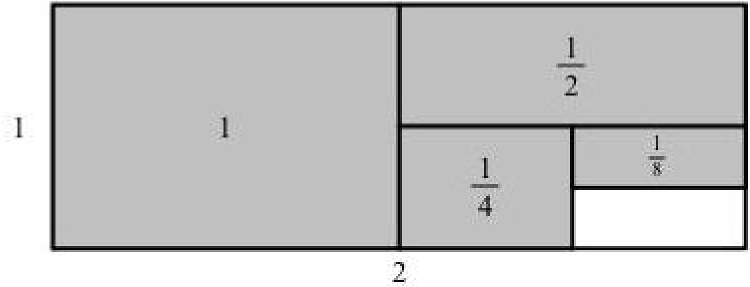
A geometry problem



Continuing in this way, the total shaded area is

$$1 + \frac{1}{2} + \frac{1}{4} = \frac{7}{4}.$$

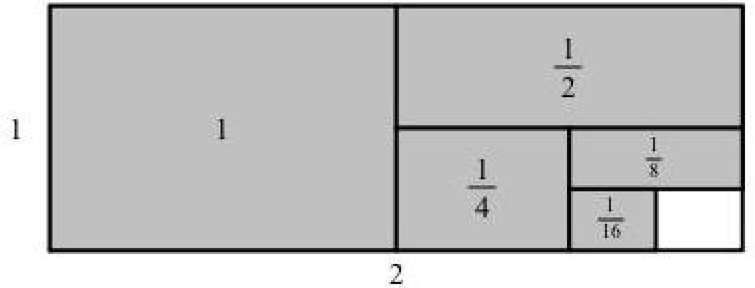
A geometry problem



And by now you get the point,

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{15}{8}.$$

A geometry problem

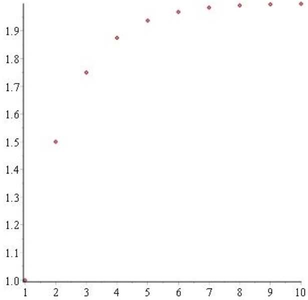


This slide is just gratuitous,

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{31}{16}.$$

A geometry problem

Suppose this pattern continues. We plot the first 10 *partial sums*.



This confirms our intuition that the sums are approaching 2. Will they ever *reach 2*?

Sequences

A sequence is an infinite list of numbers, usually denoted by a variable and an indexing: $a_1, a_2, a_3, a_4, \ldots$

Some sequences are completely random. Others are given by an explicit formula.

Examples.

- $1, -1, 1, -1, 1, -1, \ldots$
- $1, 3, 5, 7, 9, 11, 13, \ldots$
- The sequence $\{a_n\}$ where $a_n = 1 + 2n$
- $1, 1, 2, 3, 5, 8, 13, \ldots$

Sequences

A sequence is **arithmetic** if there is a common difference between successive terms.

For example, the following sequence is arithmetic

$$1, 5, 9, 13, 17, \ldots$$

In general, we can write an arithmetic sequence as

$$a_1, a_1 + d, a_1 + 2d, a_1 + 3d, \ldots$$

where d is the common difference.

Sequences

Given an arithmetic sequence, what is the sum of the first n terms?

$$\begin{aligned}
s_1 &= a_1 \\
s_2 &= a_1 + (a_1 + d) = 2a_1 + d \\
s_3 &= (2a_1 + d) + (a_1 + 2d) = 3a_1 + 3d \\
s_4 &= (3a_1 + 3d) + (a_1 + 3d) = 4a_1 + 6d \\
&\vdots \\
s_n &= na_1 + \frac{n(n-1)d}{2}
\end{aligned}$$

Arithmetic sequence example

Given the arithmetic sequence: $1, 5, 9, 13, 17, \ldots$, the common difference is 4.

The sum of the first 10 terms is

$$s_{10} = (10)(1) + \frac{(10)(9)(4)}{2} = 10 + 180 = 190.$$

Series

How do we define an *infinite sum*?

We can at least make an educated guess for the following:

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 2$$

Geometric sums

A sequence of the form $r^0, r^1, r^2, r^3, \ldots$ is called **geometric**.

What is the sum of the first k terms in a geometric sum?

$$
\begin{aligned}
s_k &= r^0 + r^1 + r^2 + r^3 + \cdots + r^{k-1} \\
rs_k &= r^1 + r^2 + r^3 + \cdots + r^{k-1} + r^k
\end{aligned}
$$

Then

$$
rs_k - s_k = r^k - r^0 \qquad \Rightarrow s_k(r-1) = r^k - r^0.
$$

Thus

$$
s_k = \frac{r^k - 1}{r - 1}
$$

Doubling pennies

Suppose you are approached with the following proposition:

- Take \$10,000,000 immediately or
- Take 1 cent today, 2 cents tomorrow, 4 cents the next day, and so on for a month (31 days).

Which option is better?

The second option is a geometric series with $r = 2$, so we have the sum

$$1 + 2^1 + 2^2 + 2^3 + \cdots + 2^{30} = \frac{2^{31} - 1}{2 - 1} = 2,147,483,647 \text{ cents.}$$

Series

A *series* is an infinite sum

$$a_1 + a_2 + a_3 + a_4 + \cdots .$$

We will often write a series in *summation notation*

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + a_4 + \cdots$$

Geometric sums

We call a series of the form $\sum_{n=1}^{\infty} r^{n-1}$ a *geometric series*.

Set $r = \frac{1}{2}$ gives our series from before.

Set $r = -\frac{1}{3}$. The geometric series is

$$\sum_{n=1}^{\infty} \left(-\frac{1}{3}\right)^{n-1} = \left(-\frac{1}{3}\right)^{0} + \left(-\frac{1}{3}\right)^{1} + \left(-\frac{1}{3}\right)^{2} + \left(-\frac{1}{3}\right)^{3} + \cdots$$
$$= 1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \cdots$$

Can we guess an upper bound for the sum of this series? What about a lower bound?

Geometric Series

If $-1 < r < 1$, then we say the geometric series $\sum_{n=1}^{\infty} r^{n-1}$ **converges** and the sum is $\frac{1}{1-r}$.

If $r \geq 1$ or $r < -1$, then we say the series **diverges**.

The series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$ is geometric with $r = \frac{1}{2}$.
Using the geometric series formula, we have

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{1 - \frac{1}{2}} = \frac{1}{1/2} = 2.$$

Geometric Series

The series $\sum_{n=1}^{\infty}\left(-\frac{1}{3}\right)^{n-1}$ is geometric and the sum is

$$\frac{1}{1-\left(-\frac{1}{3}\right)}=\frac{1}{4/3}=\frac{3}{4}.$$

Geometric Series

So then, why is $0.\overline{9} = 1$?

Notice that we can write $0.\overline{9}$ as

$$\frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \cdots = \frac{9}{10}\left(1 + \frac{1}{10} + \frac{1}{100} + \frac{1}{1000} + \cdots\right).$$

What is in parenthesis is a geometric series with $r = \frac{1}{10}$.

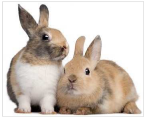
Hence,

$$0.\overline{9} = \frac{9}{10} \cdot \frac{1}{1 - \frac{1}{10}} = \frac{9}{10} \cdot \frac{1}{\frac{9}{10}} = 1.$$

Fibonacci Numbers

Breeding (magic) bunnies

Congratulations! You are now the owner of a farm of (magic) bunnies. Your farm only has one pair of bunnies right now.

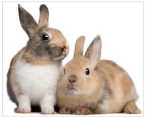


These bunnies aren't old enough to reproduce yet. After 2 months they will produce another pair of bunnies.

We'll count bunnies by pairs each month. This is the start, so we say $P_0 = 1$.

Month 1

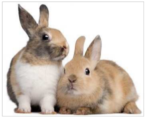
You have one pair of bunnies. Aren't they cute?



After 1 month, we still have $P_1 = 1$.

Month 2

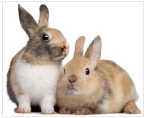
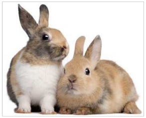
Your bunnies made new bunnies! You now have two pairs of bunnies.

After 2 months, we have $P_2 = 2$.

Month 3

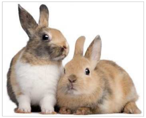
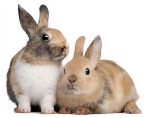
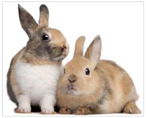
Your original bunnies had new bunnies and the bunnies from last month are all grown up!

After 3 months, we have $P_3 = 3$.

Month 4

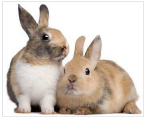
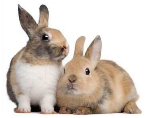
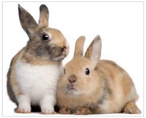
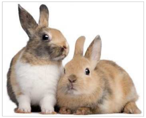
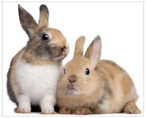
The two grown up pairs of bunnies each made a pair of bunnies.
The baby bunnies from last month are now mature.

After 4 months, we have $P_4 = 5$.

Month 5

Your bunny farm is starting to get a little out of control! The three mature pairs of bunnies each had new bunnies and the two pairs of baby bunnies are grown up.

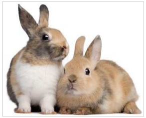
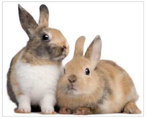
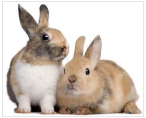
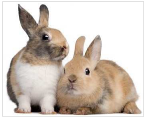
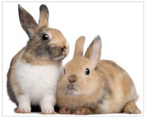
    

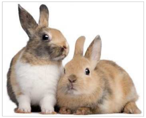
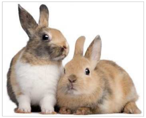
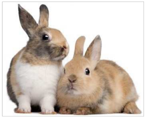
  

After 5 months, we have $P_5 = 8$.

Month 6

Local animal control would like a word with you...

After 6 months, we have $P_8 = 13$.

Month N

How many bunny pairs will we have after N months assuming we know how many bunnies we have each month prior to the Nth month?

All of the bunny pairs that are 2 months old can reproduce, so that is P_{N-2} bunnies, plus all of the bunny pairs we had the previous month, P_{N-1}.

Thus, $P_N = P_{N-2} + P_{N-1}$.

Question: Does this model accurately portray population growth?

The Fibonacci sequence

The sequence $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$ is known as the Fibonacci sequence.

The Nth Fibonacci number F_N is obtained *recursively* by adding the previous two numbers (with $F_1 = 1$ and $F_2 = 1$). So $F_N = F_{N-2} + F_{N-1}$.

Growth of Fibonacci numbers

Fibonacci numbers grow quickly.

$F_{10} = 55$

$F_{50} = 12,586,269,025$

$F_{100} = 354,224,848,179,261,915,075$

Binet's formula

The Nth Fibonacci number is given by Binet's formula:

$$F_N = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^N - \left(\frac{1-\sqrt{5}}{2}\right)^N}{\sqrt{5}}$$

A simplified version is

$$F_N = \left[\!\left[\frac{\left(\frac{1+\sqrt{5}}{2}\right)^N}{\sqrt{5}} \right]\!\right]$$

where $[\![\quad]\!]$ means 'round to the nearest integer'.

The Golden ratio

The number $\phi = \dfrac{1+\sqrt{5}}{2}$ is known as the **golden ratio**, $\phi \approx 1.618$.

The golden ratio satisfies the **golden property**: $\phi^2 = \phi + 1$. This is because ϕ is a *root* of the equation $x^2 = x + 1$. What is the other root?

Using this notation, Binet's (simplified) formula simplifies further,

$$F_N = \left[\!\left[\phi^N / \sqrt{5} \right]\!\right]$$

where $[\![\quad]\!]$ means 'round to the nearest integer'.

The Golden ratio and Fibonacci numbers

Let's consider a sequence formed by taking ratios of consecutive Fibonacci numbers:

$$\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \frac{21}{13}, \frac{34}{21}, \frac{55}{34}, \ldots$$

In decimal form, this is,

$$1, 2, 1.5, 1.666667, 1.6, 1.625, 1.615385, 1.619048, 1.617647, \ldots$$

The numbers approach the golden ratio. In fact, no matter how many decimal places you go out, at some point the sequence will match the golden ratio.

We say this sequence *coverges* to the golden ratio ϕ.

The Divine Proportion

What is the most aesthetically pleasing way to divide an object? According to the ancient Greeks, it is the following:

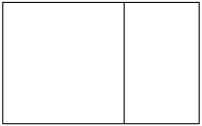
Split the object so that the ratio of the bigger piece to the smaller piece is equal to the ratio of the whole object to the bigger piece.

If we divide the object this way, then the ratio of the bigger piece to the smaller piece is the golden ratio!

The Golden Rectangle

Here is a rectangle with height 1 and base ϕ.

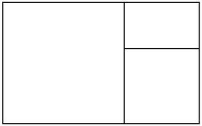
Say we divide it according to the divine proportion. This gives a square and a rectangle.



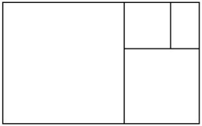
The ratio of the sides of the smaller rectangle is $1 : \phi - 1$. This is the golden ratio!

The Golden Rectangle

Let's divide it again.



The ratio of the sides of the smaller rectangle is the golden ratio.



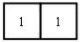
More on golden rectangles

Another way to visualize the relationship between the golden ratio and the Fibonacci sequence is through 'Fibonacci rectangles'.

We start with a rectangle of size 1 (so a square).

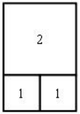
More on golden rectangles

We attach to this another rectangle of size 1.



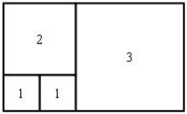
More on golden rectangles

Next we attach a rectangle of size 2.



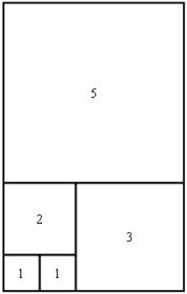
More on golden rectangles

Next we attach a rectangle of size 3.



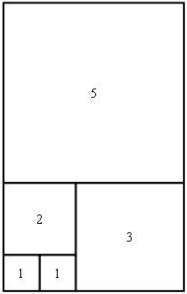
More on golden rectangles

Next we attach a rectangle of size 5.



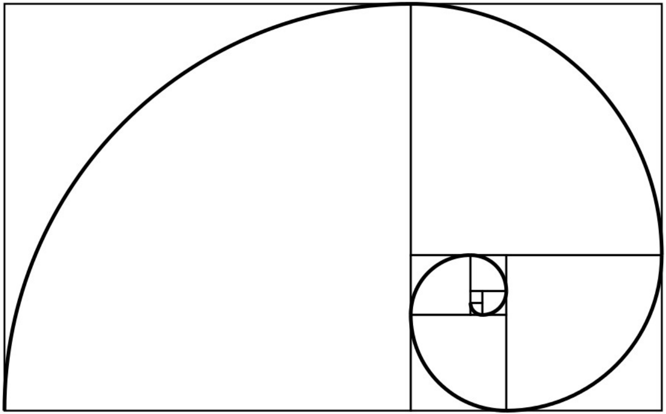
More on golden rectangles

With each iteration, the ratio of sides is closer to the golden ratio, and hence the rectangle is closer to being a golden rectangle.



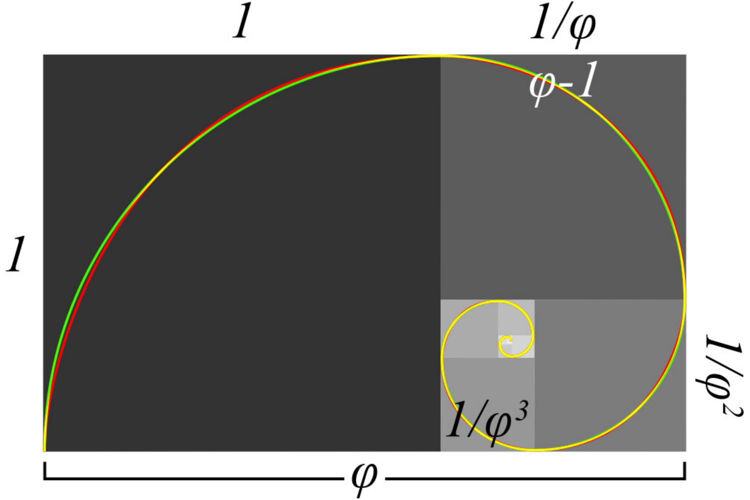
Fibonacci spiral

If we attached these Fibonacci rectangles in a certain way and drew a quarter circle through each one, the result is what is known as the **Fibonacci spiral**.



Golden spiral

The Fibonacci spiral approximates what the **Golden spiral** which is drawn through a true series of golden rectangles.



Fibonacci spiral

This type of growth appears in nature, for example in the growth of a nautilus shell. However, the growth rate may not always mirror the golden ratio but may be an approximation of it.

Population Growth

Gnomes!

This gnome just appeared in your garden and you don't know why.

$G_0 = 1$

Day 1

Today two more gnomes appeared in your garden!

$G_1 = 3$.

Day 2

Another two gnomes appeared in your garden today!

$G_2 = 5$.

Day 3

You have a gnome problem!

$G_3 = 7$.

Gnome growth

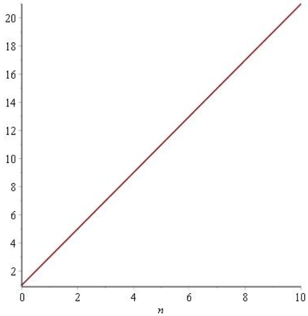
How many gnomes will there be on day 10?

The sequence of gnome counts, $G_0, G_1, G_2, G_3, \ldots$, is an arithmetic sequence with a common difference of 2.

The number of gnomes will be $2(10) + 1 = 21$.

Linear growth

Gnome growth is an example of linear growth.

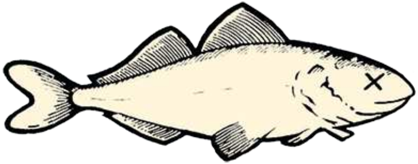


Let $a_0, a_1, a_2, \ldots$ be an arithmetic sequence with common difference d.

Then $a_n = dn + a_0$.

Dead Fish Growth

Your fish are dying at an alarming rate!



Initially 1 fish died.

Every day after that, three times as many fish die as the day before.

Dead Fish Growth

The sequence of dying fish looks like

$$1, 3, 9, 27, 81, \ldots$$

This is the geometric sequence,

$$3^0, 3^1, 3^2, 3^3, 3^4, \ldots$$

So on day n, 3^n fish will die (until you have no more fish).

Zombie Attack!

5% of students at Wake Forest have become infected with the Zombie virus and have begun infecting other students.

A zombie must feed on (and infect) an uninfected student each day in order to stay a zombie, otherwise it reverts back to normal.

Zombie Attack!

Denote the *percent* of infected students on day N by p_N (so $0 \leq p_N \leq 1$).

The growth parameter is $r = 2$. Roughly speaking, this says that in an *open* system the zombies could infect twice as many students in a day as the day before.

However, because this is a *closed* system the percentage of students infected each day is proportional to the number of students uninfected.

This is known as a Logistic Growth Model.

Logistic Growth Model

The **logistic equation**

$$p_{N+1} = r(1 - p_N)p_N$$

is a recursive formula that tells us the percentage of infected students on day $N + 1$.

Zombie Attack!

We are given that $p_0 = .05$ and $r = 2$. Using the logistic growth model, we find that

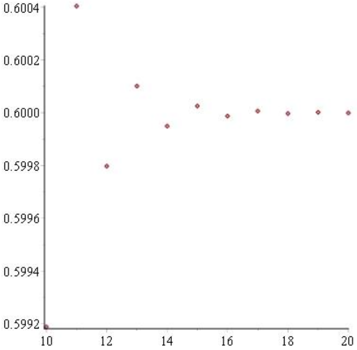
$$\begin{aligned}
p_1 &= 2(1-p_0)p_0 = 0.0950 \\
p_2 &= 2(1-p_1)p_1 = 0.17195000 \\
p_3 &= 2(1-p_2)p_2 = 0.2847663950 \\
p_4 &= 2(1-p_3)p_3 = 0.4073489906 \\
p_5 &= 2(1-p_4)p_4 = 0.4828315810 \\
p_6 &= 2(1-p_5)p_5 = 0.4994104908 \\
p_7 &= 2(1-p_6)p_6 = 0.4999993050 \\
p_8 &= 2(1-p_7)p_7 = 0.5000000000 \\
p_9 &= 2(1-p_8)p_8 = 0.5000000000 \\
p_{10} &= 2(1-p_9)p_9 = 0.5000000000
\end{aligned}$$

The values appear to have stabilized at $.5 = 50\%$.

Zombie Attack!

In the previous example, we call .5 a **stable equilibrium**.

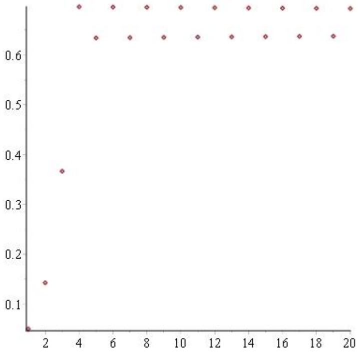
Say we instead have $r = 2.5$. In this case, the values never stabilize but they do tend towards .6.



We call this value an **attracting point**.

Zombie Attack!

What if $r = 3$?



It's harder to see, but the values are converging to $2/3$ (albeit *very* slowly).

Normal Curves

Today's Goals

- Normal curves!
- Before this we need a basic review of statistical terms. I mean basic as in underlying, not easy.
- We will learn how to retrieve statistical data from normal curves.
- As an application, we'll see how to determine the margin of error of a poll.

Statistics basics

Here's some terminology you should be familiar with:

- **Mean/Average**: For a set of N numbers, $d_1, d_2, \ldots, d_N$, the mean is given by $\mu = (d_1 + d_2 + \cdots + d_N)/N$.
- **Median**: Sort the data set from smallest to largest: $d_1, d_2, \ldots, d_N$. The median is the *middle number*. If N is odd, the median is $d_{(N+1)/2}$. If N is even, the median is the average of $d_{N/2}$ and $d_{(N/2)+1}$.
- **Mode**: The *most common number(s)*. A data set can have more than one mode. (We won't really study mode. It was just feeling left out so I put it on the slide.)
- **Range**: The difference between the highest and lowest values of the data ($R = Max - Min$).

Percentiles

The pth **percentiles** of a data set is a number X_p such that $p\%$ is smaller or equal to X_p and $(100 - p)\%$ of the data is bigger or equal to X_p.

To find the pth percentile of a *sorted* data set $d_1, d_2, \ldots, d_N$, first find the *locator* $L = (p/100)\,N$.

If L is a whole number, then $X_p = \frac{d_L + d_{L+1}}{2}$.

If L is not a whole number, then $X_p = d_{L^+}$ where L^+ is L rounded up.

Percentiles

This is Evelyn.

Evelyn is in the 40th percentile for height (40% of babies Evelyn's age weigh as much or less than she does while 60% weigh as much or more).

Quartiles

- The **first quartile** Q_1 is the 25th percentile of a data set.
- The **median** is the 50th percentile of a data set (also technically the *second quartile*).
- The **third quartile** Q_3 is the 75th percentile of a data set.
- The **fourth quartile** is d_N (the last number in the data set).

The **interquartile range (IQR)** is the difference between the third quartile and the first quartile ($IQR = Q_3 - Q_1$).

IQR tells us how spread out the middle 50% of the data values are.

Hold up!

Why aren't we doing any examples?

Because I'm not going to ask you to compute any of these things directly from a set of data. Instead, we will study visual representations of the data called *bell curves*.

But, I want you to be familiar with the terminology and how it's computed. So bear with me.

Standard deviation and variance

Standard deviation tells us how spread out a data set is *from the mean*.

Let A be the mean of a data set. For each value x in the data set, $x - A$ is the *deviation from the mean*. We want to average these values but for technical reasons we actually need to average their *squares*.

This average is called the **variance** V. The **standard deviation** is the square root of the variance, $\sigma = \sqrt{V}$.

Example

Scores (x)	Deviation (x-A)	(x-A)^2
40.00	-37.29	1390.22
41.00	-36.29	1316.65
48.00	-29.29	857.65
48.00	-29.29	857.65
70.00	-7.29	53.08
73.00	-4.29	18.37
73.00	-4.29	18.37
74.00	-3.29	10.80
77.00	-0.29	0.08
77.00	-0.29	0.08
82.00	4.71	22.22
85.00	7.71	59.51
85.00	7.71	59.51
88.00	10.71	114.80
90.00	12.71	161.65
90.00	12.71	161.65
94.00	16.71	279.37
95.00	17.71	313.80
96.00	18.71	350.22
98.00	20.71	429.08
99.00	21.71	471.51
77.29	0.00	330.78

The number 330.78 is the variance V, the average of squared devations.

The standard deviation is then

$$\sigma = \sqrt{V} \approx 18.19$$

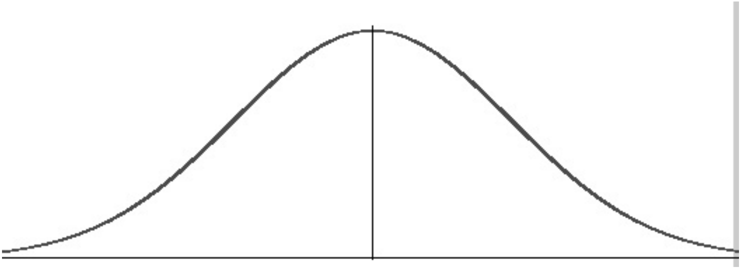
Normal curves

Say we flipped a coin 100 times? We *expect* to get heads 50 times and tails 50 times, but it's also very likely that we would not get this. (For a challenge, compute the probability of this event.)

When John Kerrich was a POW during World War II he wanted to test the probabilistic theory on coin flipping with a real life experiment. He flipped a coin 10,000 times and recorded the number of heads for each 100 trials.

What took Kerrich weeks (months?) we can do in a matter of seconds via computer software like Maple.

Bell curves



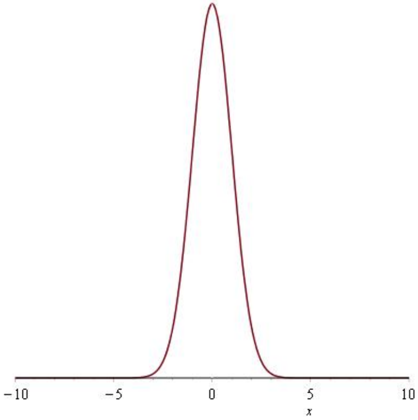
A set of data with **normal distribution** has a bar graph that is perfectly bell shaped.

Properties of normal curves

- **Symmetry**: Every normal curve has a vertical axis of symmetry.
- **Median and mean**: If a data set, then the median and mean are the same and they correspond to the point where the axis of symmetry intersects the horizontal axis.
- **Standard deviation**: The standard deviation is the horizontal distance between the mean and the **point of inflection**, where the graph changes the direction it is bending.

Normal distributions and curve

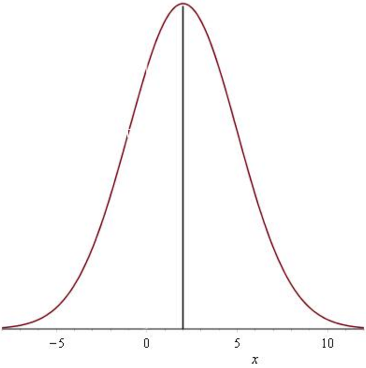
We say a distribution of data is **normal** if its bar graph is perfectly bell shaped.



This type of curve is called **normal**

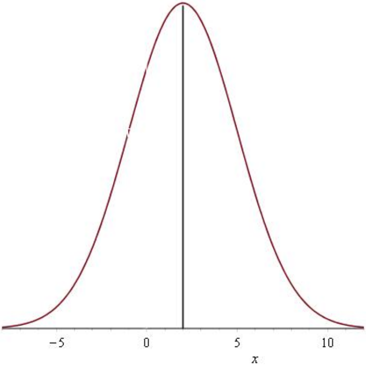
Properties of normal curves

Symmetry: Every normal curve has a vertical axis of symmetry.



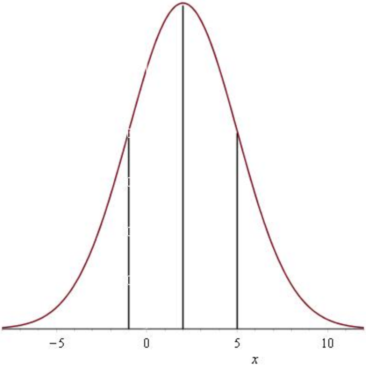
Properties of normal curves

Median and mean: If a data set is normal, then the median and mean are the same and they correspond to the point where the axis of symmetry intersects the horizontal axis.



Properties of normal curves

Standard deviation: The standard deviation is the horizontal distance between the mean and the **point of inflection**, where the graph changes the direction it is bending.



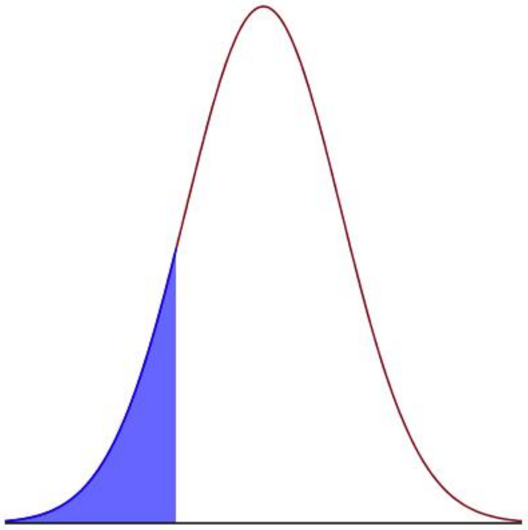
Properties of normal curves

Quartiles: The first and third quartiles can be found using the mean μ and the standard deviation σ.

$$Q_1 = \mu - (.675)\sigma \text{ and } Q_3 = \mu + (.675)\sigma.$$

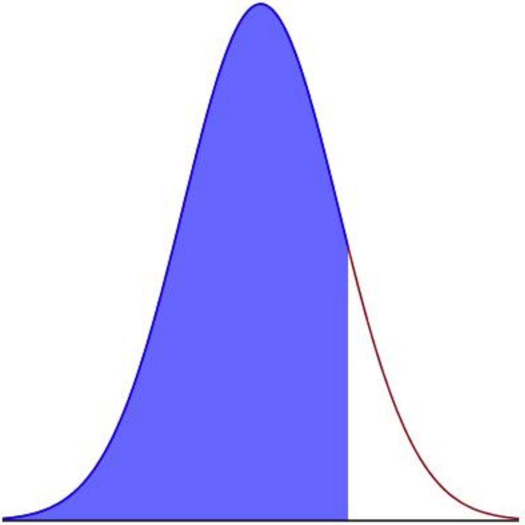
Properties of normal curves

$Q_1 = \mu - (.675)\sigma$



Properties of normal curves

$Q_3 = \mu + (.675)\sigma$



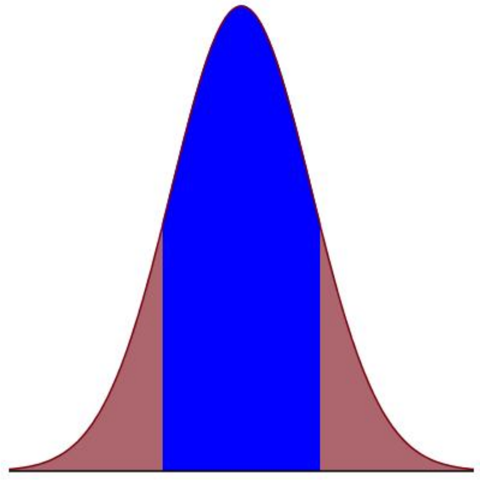
Properties of normal curves

The 68-95-99.7 Rule: In a normal data set,

- Approximately 68% of the data falls between one standard deviation of the mean ($\mu \pm \sigma$). This is the data between P_{16} and P_{84}.
- Approximately 95% of the data falls within two standard deviations of the mean ($\mu \pm 2\sigma$). This is the data between $P_{2.5}$ and $P_{97.5}$.
- Approximately 99.7% of the data falls within three standard deviations of the mean ($\mu \pm 3\sigma$). This is the data between $P_{0.15}$ and $P_{99.85}$.

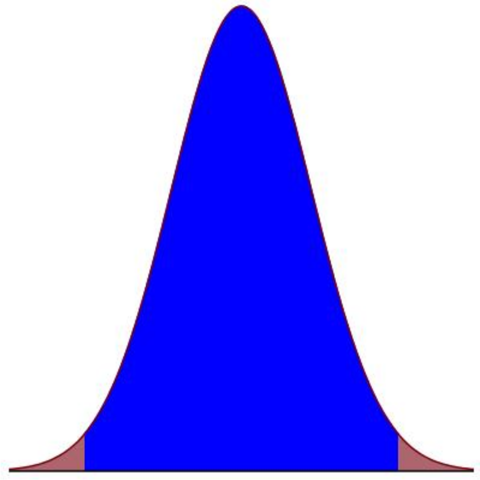
Properties of normal curves

68%



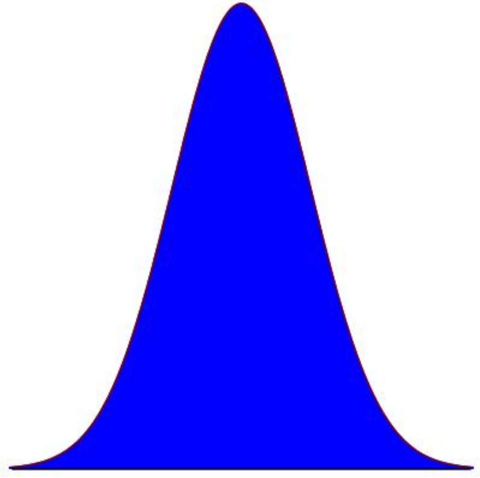
Properties of normal curves

95%



Properties of normal curves

99.7%



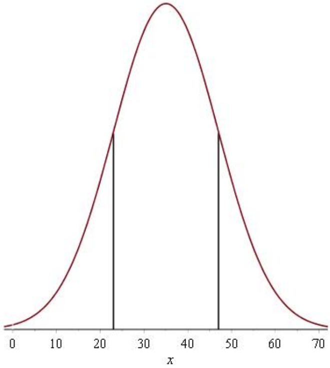
Example

Suppose we have a normal data set with mean $\mu = 500$ and standard deviation $\sigma = 150$. We have the following:

- $Q_1 = 500 - .675 \times 150 \approx 399$
- $Q_3 = 500 + .675 \times 150 \approx 601$
- Middle 68%: $P_{16} = 500 - 150 = 350$, $P_{84} = 500 + 150 = 650$.
- Middle 95%: $P_{2.5} = 500 - 2(150) = 200$, $P_{97.5} = 500 + 2(150) = 800$.
- Middle 99.7%: $P_{0.15} = 500 - 3(150) = 50$, $P_{99.85} = 500 + 3(150) = 850$.

Example

Consider a normal distribution represented by the normal curve with points of inflection at $x = 23$ and $x = 45$. Find the mean and standard deviation. Use them to compute Q_1, Q_3 and the middle 68%, 95%, and 99.7%.



Standardizing normal data

In essence, all normalized data sets are the same. They all have a mean μ and standard deviation σ. The same percentage of data is located in the same increments of σ from the mean. Thus, there is value in *standardizing normal data*.

Psychometry

This is Laura.

Laura is a *psychometrist*. She conducts psychological assessments.

Her patients are adults but their ages range from 18 and up. She uses z-values to standardize her patients' assessment scores.

Standardizing Rule

In a normal distribution with mean μ and standard deviation σ, the standardized value of a data point x is

$$z = \frac{x - \mu}{\sigma}.$$

The result of this is the **z-value** of the data point x.

Conversions

Suppose we have a normal data set with mean $\mu = 120$ and standard deviation $\sigma = 30$. If $x = 100$, then the z-value of x is

$$z = \frac{x - 120}{30} = -\frac{2}{3} \approx -.67.$$

If a z-value of some x is .5, what is x (for the data above)?

$$\begin{aligned} .5 &= \frac{x - 120}{30} \\ 15 &= x - 120 \\ 135 &= x \end{aligned}$$

Or, we could recognize that a z-value of .5 means that x is $\frac{1}{2}$ a standard deviation to the right of the mean (so $120 + 15 = 135$).

Variables

In algebra, a variable typically is a placeholder for some type of solution or set of solutions.

Given the equation $x + 3 = 10$, then the variable x represents the number 7.

Given the equation $x^2 + 5 = 21$, then x represents a member of the set of solutions $\{-4, 4\}$.

Random variables

A variable representing a random (probabilistic) event is called a **random variable**.

For example, if we toss a coin 100 times and let X represent the number of times heads comes up, then X is a random variable.

Like an algebraic variable, X represents a number between 0 and 100, but the possible values for X are not equally likely.

The probability of $X = 0$ or $X = 100$ is $(1/2)^{100}$, which is a very small number.

The probability of $X = 50$ is about 8%.

Random variable

Continuing with the example, we know that X has an approximately normal distribution with mean $\mu = 50$ and standard deviation $\sigma = 5$ (for a sufficiently large number of repetitions).

What is the (approximate) probability that X will fall between 45 and 55? This is 1 standard deviation from the mean, so the probability is approximately 68%.

The Honest-Coin Principle

We can now generalize the previous example to a trial with n tosses.

Let X be a random variable representing the number of heads in n tosses of an honest (fair) coin (assume $n \geq 30$).

Then X has an approximately normal distribution with mean $\mu = n/2$ and standard deviation $\sigma = \sqrt{n}/2$.

The Dishonest-Coin Principle

Let X be a random variable representing the number of heads in n tosses of a coin (assume $n \geq 30$), and let p denote the probability of heads on each toss of the coin.

Then X has an approximately normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$.

Note that when $p = \frac{1}{2}$ we recover the Honest-Coin Principle.

Margin of Error

In a poll conducted by Public Policy Polling before the recent Democratic primary in Missouri interviewed 839 likely voters.

Their poll found almost a tie between Hillary Clinton and Bernie Sanders.

Therefore, we can use the Honest-Coin Principle to compute the margin of error for the poll.

Margin of Error

According the Honest-Coin Principle, we have

$$\mu = \frac{839}{2} = 419.5 \text{ and } \sigma = \frac{\sqrt{839}}{2} = 14.48.$$

The standard deviation σ is approximately 1.72% of the sample.

This means that the pollsters could assume with 95% confidence that either candidate would get between $(50 \pm 2(1.72))\%$ of the vote. That is, between 46.55% and 53.45%.

The value 2σ is called the **margin of error**.

Margin of Error

On the other hand, in a poll conducted by Public Policy Polling before the recent Democratic primary in North Carolina interviewed 747 likely voters.

Their poll found Hillary Clinton with 60% support and Bernie Sanders with 40%.

Therefore, we can use the Dishonest-Coin Principle to compute the margin of error for the poll.

Margin of Error

According the Dishonest-Coin Principle, we have

$$\mu = 747 * .6 = 448.2 \text{ and } \sigma = \sqrt{747 * .6 * .4} = 13.39.$$

The standard deviation σ is approximately 1.79% of the sample.

This means that the pollsters could assume with 95% confidence that Clinton candidate would get between $(60 \pm 2(1.79))\%$ of the vote. That is, between 56.42% and 63.58%.

The margin of error in this example is $2\sigma = 3.58\%$.

Hypothesis Testing

The Null Hypothesis

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so."

- from *The Big Short* (not Mark Twain)

"The most difficult subjects can be explained to the most slow-witted man if he has not formed any idea of them already; but the simplest thing cannot be made clear to the most intelligent man if he is firmly persuaded that he knows already, without a shadow of a doubt, what is laid before him."

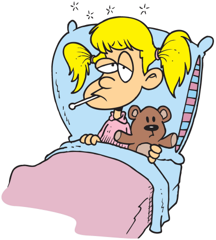
- Leo Tolstoy

Today's Goals

- Define hypothesis testing rigorously.
- Learn how to properly state the null hypothesis.
- Given data, we will check whether the null hypothesis may be rejected.

Example (from Ellenberg)

Say we have a study of 100 patients suffering from a certain disease. These patients have a 10% chance of dying with no intervention.



50 patients will be given a new experimental drug and the other 50 will be given a placebo.

Example (from Ellenberg)

Our *hypothesis* is that the drug is more effective than the placebo in fighting the disease. That is, we expect fewer of the drug patients will die than the placebo patients.

If we observe this event, can we claim success? **No!**

Even if the drug is totally useless, it is still *not unlikely* that fewer drug patients will die.

- 13.3% chance equally many placebo and drug patients die
- 43.3% chance fewer drug patients than placebo patients die
- 43.3% chance fewer placebo patients than drug patients die

Example (from Ellenberg)

What if *none* of the drug patients die?

Even though each patient has a 90% probability of survival, the odds of this happening by pure chance is

$$(.9)^{50} \approx 0.00515,$$

or about .5%. That is *very unlikely*.

This is closer to the right way to frame the question.

Null Hypothesis

The **null hypothesis** is the statement that there is *no relationship* between two phenomena.

What is the null hypothesis in our example?

There is no relationship between the drug and survival rates.

Suppose this were true in our example. If no drug patients died, then it would be very unlikely that the null hypothesis is true and we must reject it.

Thus, we conclude that there likely is a relationship between the drug and survival rates among patients.

Null Hypothesis

What if one drug patient died?

The probability is

$$(.1)^1 \cdot (.9)^{49} \cdot 50 \approx 0.0286,$$

or about 2.9%. This is still unlikely.

In this scenario we are still safe in rejecting the null hypothesis.

Aside on Probability

On the last side, we computed the probability of one patient dying as

$$(.1)^1 \cdot (.9)^{49} \cdot 50 \approx 0.0286.$$

Where did this formula come from?

There is a 10% chance of any one patient dying and a 90% chance of any one patient living. We want the situation where exactly one patient dies and the other 49 live.

But how many ways can this happen? There are 50 patients so if any one of them dies we are in the right situation. Hence, we multiply our probability by 50.

Null Hypothesis

What if two drug patients died?

The probability is

$$(.1)^2 \cdot (.9)^{48} \cdot 1225 \approx 0.0779,$$

or about 7.8%. This is too high to reject the null hypothesis.

What is the correct threshold to have for rejecting the null hypothesis? The now commonly (but not universally) accepted threshold is 5%, originally set by R.A. Fisher.

(Another) Aside on Probability

On the last side, we computed the probability of two patients dying as

$$(.1)^2 \cdot (.9)^{48} \cdot 1225 \approx 0.0779,$$

Where did this formula come from?

Again, we want the situation where two patients die and the other 48 live. But which two? Any pair will give us a valid situation.

The 1225 comes from asking the following question: How many ways can we choose a set of 2 from a set of 50?

There's a handy little formula for answering this very question.

Binomial Coefficients

For a positive number k, define k **factorial** to be

$$k! = k \cdot (k-1) \cdot (k-2) \cdots 2 \cdot 1.$$

By convention we set $0! = 1$.

The number of ways of choosing a set of r elements from a set of n elements is

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Binomial Coefficients (example)

We want to choose a set of 2 from a set of 50. This is

$$_{50}C_2 = \binom{50}{2} = \frac{50!}{2!(48)!} = \frac{50 \cdot 49}{2} = 1225.$$

Null Hypothesis

In statistical hypothesis testing, there are always two hypotheses. The hypothesis to be tested is called the null hypothesis and given the symbol H_0.

The null hypothesis states that there is no difference between a hypothesized population mean and a sample mean. It is the status quo hypothesis.

We test the null hypothesis against an alternative hypothesis, which is given the symbol H_a. The alternative hypothesis is often the hypothesis that you believe yourself! It includes the outcomes not covered by the null hypothesis.

Examples

If we were to test the hypothesis that college freshmen study 20 hours per week, we would express our null hypothesis as:

$$H_0 : \mu = 20.$$

In this example, our alternative hypothesis would express that freshmen do not study 20 hours per week:

$$H_a : \mu \neq 20.$$

Examples

Suppose instead that our hypothesis was that college freshman study at least 20 hours per week. Then we would express our null hypothesis as:

$$H_0 : \mu \geq 20.$$

The alternate hypothesis would then be that freshman study fewer that 20 hours per week:

$$H_a : \mu < 20.$$

Example

We have a medicine that is being manufactured and each pill is supposed to have 14 milligrams of the active ingredient. What are our null and alternative hypotheses?

$$H_0 : \mu = 14$$
$$H_a : \mu \neq 14.$$

Our null hypothesis states that the population has a mean equal to 14 milligrams. Our alternative hypothesis states that the population has a mean that is different than 14 milligrams.

The Null Hypothesis Significance Test

1. Run an experiment.
2. Suppose the null hypothesis is true, and let p be the probability (under that hypothesis) of getting results as extreme as those observed.
3. The number p is called the p-value. If it is very small ($p < 0.05$), then the result is *statistically significant*.

Example

Suppose we think that a coin is biased towards heads.

$$H_0 : \text{The coin is fair}$$
$$H_a : \text{The coin is biased towards heads}$$

In our experiment, the coin is flipped 10 times and 8 heads are observed.

To compute the p-value of this outcome, we need the probability that *at least* 8 heads are flipped.

Example

$$P[8 heads] = {}_{10}C_8 \cdot \left(\frac{1}{2}\right)^8 \cdot \left(\frac{1}{2}\right)^2 \approx 0.0439$$

$$P[9 heads] = {}_{10}C_9 \cdot \left(\frac{1}{2}\right)^9 \cdot \left(\frac{1}{2}\right)^1 \approx 0.0098$$

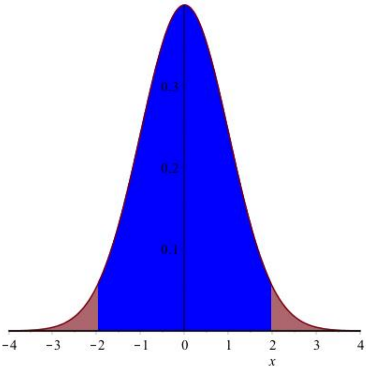
$$P[10 heads] = {}_{10}C_{10} \cdot \left(\frac{1}{2}\right)^{10} \cdot \left(\frac{1}{2}\right)^0 \approx 0.00098.$$

The probability that at least 8 heads are flipped is

$$P[8 heads] + P[9 heads] + P[10 heads] \approx 0.055.$$

Thus, we *cannot* reject the null hypothesis.

Normal distributions and the null hypothesis



Suppose our experimental data is normal (or approximately normal) with mean μ and standard deviation σ.

If the probability *p* of the null hypothesis lies *outside* of the 95% confidence interval then the result is significantly significant. That is, *p* is *more than* two standard deviations away from the mean.

This is known as the **two-tailed test**.

Normal distributions and the null hypothesis

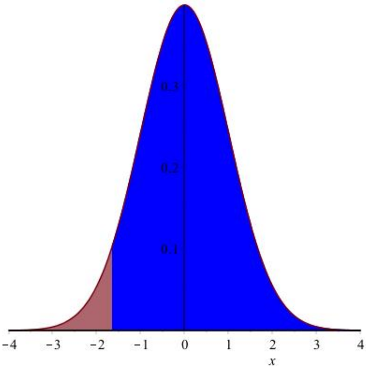
Quick Accuracy Note: On the last slide we used the 68-95-99.7 rule that says that *approximately* 95% of the data lies within two standard deviations of the mean.

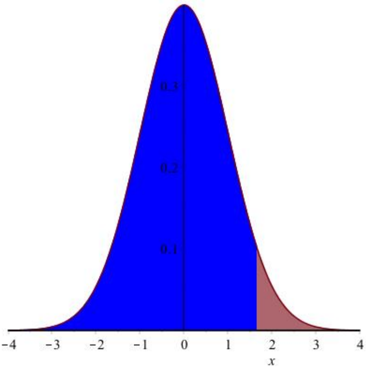
In practice, this level of approximation is not quite good enough. It is more accurate to say that approximately 95% of the data lies within 1.96 standard deviations of the mean.

This is still an approximation (but a better one).

Normal distributions and the null hypothesis

In the **one-tailed test**, we look for a *p*-value that is strictly greater or less than the mean. The cutoff for $\alpha = 0.05$ is ± 1.645.





Example

Suppose we think that a coin is biased towards heads.

Our null hypothesis is that the coin is fair. That is, the null hypothesis is a sequence of Bernoulli trials with probability 0.5, yielding a random variable X which is 1 for heads and 0 for tails, and a common test statistic is the sample mean (of the number of heads) $\overline{x}$.

A data set of five heads (HHHHH) with sample mean of 1 has a $(1/2)^5 = 1/32 = 0.03125 \approx 0.03$ chance of occurring, and thus would would be significant (rejecting the null hypothesis) if using 0.05 as the cutoff.

Example

However, if testing for whether the coin is biased towards heads or tails, a two-tailed test would be used.

A data set of five heads (sample mean 1) is as extreme as a data set of five tails (sample mean 0), so the p-value would be $2/32 = 0.0625 \approx 0.06$ and this would not be significant (not rejecting the null hypothesis) if using 0.05 as the cutoff.

Z-test

The Z-test is an effective way of checking the null hypothesis assuming a normal distribution of data. This will look relatively familiar from our discussion of *z*-scores.

Once computed, we can use a table to determine the probability of observing values less than or greater than (or both) this statistic.

If this probability is less than the threshold (usually $\alpha = 0.05$ or $\alpha = 0.01$), then we can reject the null hypothesis.

Z-test

Consider a population normally distributed with mean μ and standard deviation σ. Suppose we poll a **simple random sample (SRS)** of n respondents within this population.

- Let $\overline{x}$ be the sample average.
- Let μ_0 be the value corresponding to the null hypothesis.
- The **standard error** is $SE = \sigma/\sqrt{n}$.
- The *z*-**statistic** is

$$z = \frac{\overline{x} - \mu_0}{SE}$$

Example

Suppose that in a particular geographic region, the mean and standard deviation of scores on a reading test are 100 points, and 12 points, respectively.

A sample of 55 students in a particular school received a mean score of 96. That is, Let $\overline{x} = 96$ and $\sigma = 12$. We can ask whether this mean score is significantly lower than the regional means.

H_0 : The sample of students are average compared to the region.

H_a : The sample of students are below average compared to the region.

Example

The standard error is $SE = \dfrac{\sigma}{\sqrt{n}} = \dfrac{12}{\sqrt{55}} = 1.62$.

The z-statistic is $z = \dfrac{\overline{x} - \mu_0}{SE} = \dfrac{96 - 100}{1.62} = -2.47$.

Using an appropriate table, we can look up that the probability of observing a standard normal value below -2.47 is approximately 0.0068 (one-sided p-value). The two-sided p-value is $2 \cdot 0.0068 \approx 0.014$. This is statistically significant.

The Z-test tells us that the 55 students of interest have an unusually low mean test score compared to most simple random samples of similar size from the population of test-takers.

Example

The mean weight in men is estimated to be 191 pounds with standard deviation $\sigma = 25.6$. Assume we test an SRS of 100 men and find the sample mean in $\overline{x} = 197.1$. Researchers want to know if this estimate is correct.

$$H_0 : \mu = 191$$
$$H_a : \mu > 191.$$

Example

The standard error is $SE = \frac{\sigma}{\sqrt{n}} = \frac{25.6}{\sqrt{100}} = 2.56$.

The z-statistic is $z = \frac{\overline{x} - \mu_0}{SE} = \frac{197.1 - 191}{2.56} = 2.38$.

Using the one-tailed test, we find that the probability that this occurs (by chance) is 0.0087 (or about .87%). Hence, at $\alpha = 0.05$ significance we can reject the null hypothesis. What about $\alpha = 0.005$?

Errors

There are two potential errors that can occur when running significance tests.

- Type I error: Rejecting the null hypothesis when it is true (saying false when true). (Usually the more serious error.)
- Type II error: Failing to reject the null hypothesis when it is false (saying true when false).

Errors

In the jury trial there are two types of errors:

- The defendant is innocent but the jury finds the person guilty.
- The defendant is guilty but the jury declares the person to be innocent.

In our system of justice, the first error is considered more serious than the second error.

	Defendant is Innocent	Defendant is Guilty
Verdict: Innocent	Correct Decision	Type II error
Verdict: Guilty	Type I error	Correct Decision

Bayesian Inference

Today's Goals

- Define Bayesian Inference.
- Learn how to determine probabilities based on observed data.

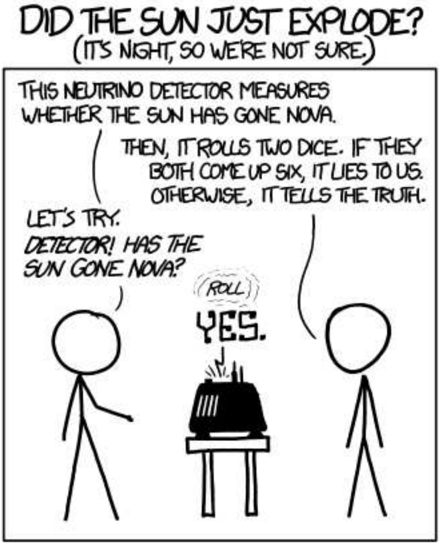
Frequentists vs Bayesians

A **frequentist** thinks of unknown parameters as fixed.

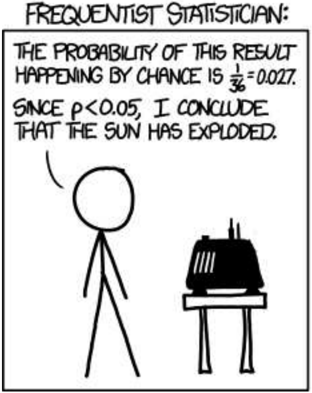
A **Bayesian** thinks of parameters as random, and thus having distributions (just like the data).

A Bayesian writes down a prior guess for θ and combines it with the likelihood for the observed data Y to obtain the posterior distribution of θ. All statistical inferences then follow from summarizing the posterior.

Frequentists vs Bayesians



Frequentists vs Bayesians



Frequentists vs Bayesians

Frequentists vs Bayesians

What is wrong with the frequentist's conclusion?

The frequentist is a victim of the **base rate fallacy**. That is, the frequentist has dismissed an unlikely explanation even though the alternative is even less likely.

A star has never spontaneously exploded, thus the likelihood of that even happening is near zero.

On the other hand, the probability that the machine is lying is $1/36 \approx 3\%$. That is unlikely but not impossible.

Conditional Probability

If A is an event, then $P(E)$ denotes the probability that this event occurs.

If A and B are both events, then $P(A|B)$ denotes the probability that A occurs *given that* B occurs. This is known as a **conditional probability**.

Conditional Probability

This is DJ LeMahieu.

In 2016, his batting average was .348. In terms of probability, $P(H) = .348$ where H is the event that he gets a hit.

In 2016, his batting average *with a runner in scoring position* was .316. In terms of probability, $P(H|S) = .316$ where S is the event that there is a runner in scoring position.

Example

Suppose we have a standard deck of 52 cards. Say we draw 3 cards. What is the probability that these cards will all be clubs?

The probability that the first card is clubs is $13/52$.

The probability that the second card is clubs *given that* the first card is clubs is $12/51$.

The probability that the third card is clubs *given that* the first two cards were clubs is $11/50$.

The total probability is then

$$\frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} = \frac{11}{850} \approx 0.01294117647.$$

Bayes' Theorem

Bayes' Theorem gives a way for computing conditional probabilities.

If A and B are two events, then Bayes' Theorem says

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

The critical point about **Bayesian inference** is that it provides a principled way of combining new evidence with prior beliefs, through the application of Bayes' rule.

Example

Suppose that you are given two drawers. You cannot see the contents of the drawers, but you are told that one drawer contains two gold coins and the other drawer contains one gold coin and one silver coin. If someone pulls a coin at random out of drawer A and it turns out to be gold, what is the probability that drawer A is the drawer with two gold coins?

Example

A: the event that Drawer A has two gold coins ($P(A) = 0.5$).

B: the event that the person chooses a gold coin out of the four coins ($P(B) = 0.75$).

$P(B|A)$: the conditional probability of choosing a gold coin from Drawer A if it has two gold coins ($P(B|A) = 1.0$).

Bayes Theorem can then be used to compute the conditional probability that Drawer A has the two gold coins *given that* the person chose a gold coin from Drawer A.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{(1.0)(0.5)}{0.75} = \frac{2}{3}.$$

Probability

Today's Goals

- Discuss terminology for our unit on Expected Value.
- In particular, we will study how mathematicians measure risk.
- We will look at several examples of probability experiments and games where one can gain an advantage by knowing the odds ahead of time.

Terminology

- **Random Experiment**: An activity or process whose outcome cannot be predicted ahead of time.
- **Sample Space**: The set of all possible outcomes of the experiment. (We will use S to represent this set and N to denote the number of possible outcomes.)

In all of our examples, we will assume N is a finite number, though it is certainly possible to consider examples where N is infinite.

Tossing a coin

Suppose we toss a coin. There are two possible outcomes: heads (H) or tails (T). The sample space for this experiment is

$$S = \{H, T\} \text{ and } N = 2.$$

Now suppose we toss a coin twice. How many possible outcomes are there?

$$S = \{HH, HT, TH, TT\} \text{ and } N = 4.$$

Suppose we toss a coin k times. What is the size of the sample space?

Probability

We will be interested in determining the *probability* or likelihood of a certain *event*.

The following questions don't need exact answers yet, but intuitively we might have an idea of the answers.

- If we toss a coin once, what is the probability that it lands on heads?
- If we toss a coin twice, what is the probability that it lands on heads at least once?

Rolling dice

This is like the coin toss example but with more options.Suppose we roll a six-sided die. The sample space looks like $S = \{1, 2, 3, 4, 5, 6\}$.

Now suppose we roll two six-sided die. We could represent the sample space in pairs: $S = \{(1, 1), (1, 2), (1, 3), \cdots\}$. But there are a lot of options! (How many?)

Instead, we might say we are interested in the *sum* of the two numbers, in which case our sample space will be $S = \{1, 2, 3, \ldots, 12\}$.

Now the questions get harder. What is the probability of rolling 12? What about 11?

Blackjack

The rules: You are dealt two cards and the dealer is dealt two cards, one face up. Cards 2-7 are worth their number, face cards are worth 10, an ace is worth 1 or 11.

You have the option to **stand** (keep the cards you are dealt) or **hit** (get another card). (There are other possibilities such as doubling down and splitting.)

Basic blackjack strategy involves knowing what events are possible if you were to hit. When playing at a casino, using basic strategy gives you are 49% chance of winning *in the long run*. When card counting is used to also take into account the cards left in the deck, your chance increases to about 52% *in the long run.*

Events

- An **event** is a subset of the sample space. (Recall: How many subsets are there of a set of size N, including the empty set?)
- The empty set $\{ \}$ is called the **impossible event**.
- The set itself is called the **certain event**.
- The events containing a single outcome (sets of size 1) are called **simple events**.

What are all of the events from the experiment when we flipped a coin twice?

Probability

Now that we have the terminology, we can talk about probability a bit more concretely.

A **probability assignment** is a function that assigns to each event E a number between 0 and 1. We denote this $\mathrm{Pr}(E)$.

The impossible event should always be assigned a probability of 0 ($\mathrm{Pr}(\{\ \}) = 0$) and the certain event should always be assigned a probability of 1 ($\mathrm{Pr}(S) = 1$).

Since our sample spaces are finite, probability assignments are determined by assigning a probability to just the simple events.

The probability of an event is then obtained by adding the probabilities of the individual outcomes that make up that event.

A valid probability assignment satisfies two requirements:

- All probabilities are numbers between 0 and 1.
- The sum of the probabilities of the simple events is 1.

Probability

Suppose we have an experiment (like tossing a coin) where each *simple event* has the same probability.

Then each simple event occurs with probability $1/N$ where N is the size of the sample space.

This means that the probability of an event E of size k occurs with probability k/N (so $\Pr(E) = k/N$).

Such a sample space is called **equiprobable**.

Probability

Suppose we roll two six-sided dice. This is an equiprobable space *if* we consider the sample space as pairs of numbers 1-6.

In this case, the probability that we roll a 2 on the first die and a 3 on the second die is $1/36$.

However, if we write our sample space using sums then this is no longer equiprobable. That is, we are *more likely* to roll a 5 than a 4. Why is this?

Probability

What is the probability of rolling a sum of 5? How do we compute it?

In order to figure this out we should consider the number of ways that a 5 could occur:

$$(1,4),(2,3),(3,2),(4,1).$$

Hence, the probability of rolling a 5 is $4/36 = 1/9$.

Shooting Free Throws

This is Shaquille O'Neal, or Shaq.

Shaq was a 15-time NBA all-star and notoriously bad free throw shooter. He was so bad that other teams would employ the hack-a-shaq strategy to force him to shoot free throws.

During his career, he made approximately 52% of his free throws.

Shooting Free Throws

Suppose Shaq is fouled (hack-a-shaq!) and gets a free throw. The different outcomes are a success (s) or a failure (f).

What is the solution space? $S = \{s, f\}$.

This is not an equiprobable space. What are the probabilities of the events $\{s\}$ and $\{f\}$? $\Pr(\{s\}) = .52$ and $\Pr(\{f\}) = 1 - .52 = .48$.

In this case, the event $\{f\}$ is the **complement** of the event $\{s\}$ because $\Pr(\{f\}) = 1 - \Pr(\{s\})$. Said another way, they are complements because the sum of their probabilities is 1.

Shooting Free Throws

Was hack-a-shaq a good strategy? On first glance it's not because he still made free throws more often than he missed. But what about near the end of the game when a player might get 2 free throws.

Suppose Shaq is fouled and gets 2 free throws. What is the solution space? $S = \{ss, sf, fs, ff\}$.

What is the probability he makes both?

Shooting Free Throws

We will ignore any psychological pressure associated with this problem and assume that each free throw is independent.

The **multiplication principle for independent events** says that if E and F are independent (one event occurring does not effect the probability of the other) then

$$\Pr(E \text{ and } F) = \Pr(E) \times \Pr(F).$$

This means that the probability of Shaq making *both* free throws is $(.52) \times (.52) \approx .27$.

Shooting Free Throws

Let's continue with this and compute the other probabilities in the solution space for two free throws.

- $\Pr(\{sf\}) = (.52) \times (.48) \approx .25$
- $\Pr(\{fs\}) = (.48) \times (.52) \approx .25$
- $\Pr(\{ff\}) = (.48) \times (.48) \approx .23$

What is the probability that Shaq makes *at least* one free throw? Here, the even is $E = \{ss, sf, fs\}$, so we could get the probability by addition,

$$\Pr(E) = \Pr(\{sf\}) + \Pr(\{fs\}) + \Pr(\{ff\}) \approx .77.$$

Or we could recognize that this is the complement of the event $F = \{ff\}$ (he makes no free throws). In which case

$$\Pr(E) = 1 - \Pr(F) \approx .77$$

Shooting Free Throws

Here's a problem to think about: Say Shaq gets a 1-and-1, that is, he gets to shoot a free throw and if he makes it he can shoot another one.

(Acknowledgement: This isn't really a thing in the NBA but it's still an interesting problem.)

What is the solution space? Compute the probabilities for the different events. How is this problem different from the last one?

Shooting Free Throws

Another problem to think about, but harder: Shaq's average of making a shot without being fouled was about 58% in his career. Assuming such a shot was worth 2 points, was hack-a-shaq a good strategy in the long run?

What are some factors not taken into account here? How could we make this analysis more precise?

Expected Value

Today's Goals

- Discuss terminology for our unit on Expected Value.
- In particular, we will study how mathematicians measure risk.
- We will look at several examples of probability experiments and games where one can gain an advantage by knowing the odds ahead of time.

Terminology

- **Random Experiment**: An activity or process whose outcome cannot be predicted ahead of time.
- **Sample Space**: The set of all possible outcomes of the experiment. (We will use S to represent this set and N to denote the number of possible outcomes.)

In all of our examples, we will assume N is a finite number, though it is certainly possible to consider examples where N is infinite.

Events

- An **event** is a subset of the sample space. (Recall: How many subsets are there of a set of size N, including the empty set?)
- The empty set $\{ \ \}$ is called the **impossible event**.
- The set itself is called the **certain event**.
- The events containing a single outcome (sets of size 1) are called **simple events**.

The Multiplication Rule

In a given experiment, how many events do we expect? This depends somewhat on *how* you want to count.

Multiplication Rule: If there are m different ways to do X and n different ways to do Y, then X and Y together (in that order) can be done in $m \times n$ different ways.

Where have we seen this before?

Example: Frozen Custard

This is Kopps.

Kopps does not serve ice cream.

It serves **frozen custard** (and jumbo burgers).

Example: Frozen Custard

Every day at Kopps you can get chocolate or vanilla custard, plus there are two *flavor of the day* options.
Today the options are Godiva Chocolate Cherry and Chocolate Lover's Banana Split.

You can choose between a cup, a cake cone, or a waffle cone.

Example: Frozen Custard

How many combinations are there of custard and container?

$$(\# \text{ of custard options }) \times (\# \text{ of container options }) = 4 \times 3 = 12.$$

Example: Frozen Custard

Suppose you want a double with two different flavors. How many flavor combinations are there?

There are 4 options for the first choice and 3 options for the second choice. So $4 \times 3 = 12$ options, right?

Chocolate and Vanilla is the same as Vanilla and Chocolate, so we have double counted. This means that there are really $12/2 = 6$ choices.

Unless order matters to you. Then there really are 12 choices.

Permutations and Combinations

Permutation: An ordered selection of r objects chosen from a set of n objects.

$$_nP_r = \frac{n!}{(n-r)!}$$

Combination: An unordered selection of r objects chosen from a set of n objects.

$$_nC_r = \frac{n!}{r!(n-r)!}$$

Permutations and Combinations

Given 10 songs, how many 6 song playlists can we construct?

If order matters

$$_{10}P_6 = \frac{10!}{(10-6)!} = \frac{10!}{4!} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 151,200.$$

If order doesn't matter

$$_{10}C_6 = \frac{10!}{6!(10-6)!} = \frac{10!}{6!4!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210.$$

Probability

A **probability assignment** is a function that assigns to each event E a number between 0 and 1. We denote this $\Pr(E)$.

A valid probability assignment satisfies two requirements:

- All probabilities are numbers between 0 and 1.
- The sum of the probabilities of the simple events is 1.

The impossible event should always be assigned a probability of 0 ($\Pr(\{\ \}) = 0$) and the certain event should always be assigned a probability of 1 ($\Pr(S) = 1$).

Weighted average

Let X be a variable that takes the values $x_1, x_2, \dots, x_N$, and let $w_1, w_2, \dots, w_N$ denote the respective weights for these values, with $w_1 + w_2 + \cdots + w_N = 1$.

The **weighted average** for X is given by

$$x_1 \cdot w_1 + \cdots + x_N \cdot w_N.$$

Example: Grades

This is Tom.

Tom spends too much time on Myspace and not enough time studying for Math 107.

Example: Grades

Here are Tom's grades:

- Classwork/Participation: 0 %
- Reflections: 90%
- Homework: 60 %
- Projects: 80 %
- Exams: 70 %

In Math 107, Classwork/Participation is 10%, Reflections are 10%, Homework is 10 %, Projects are 30%, Exams are 40%.Thus, Tom's weighted average is,

$$(0 \times .1) + (90 \times .1) + (60 \times .1) + (80 \times .3) + (70 \times .4) = 67.$$

Don't be like Tom.

Expected value

Expected value is the probabilistic version of a weighted average.

Suppose that X is a variable that takes on the numerical values (outcomes) $x_1, x_2, \ldots, x_N$ with probabilities $p_1, p_2, \ldots, p_N$, respectively.

The **expected value** (or expectation) for X is given by

$$E = x_1 \cdot p_1 + \cdots + x_N \cdot p_N.$$

That is, it is the sum of each outcome times its probability.

Example: Going for 2

When a team in the NFL scores a touchdown, that team has two options: kick an extra point or go for a two-point conversion. New rules this past season made extra points less automatic.

During the 2015 season, the Jacksonville Jaguars made 82.1% of their extra point kicks. The expected value of an extra point is

$$(1 \times .821) + (0 \times .179) = .821.$$

Jacksonville made only 20% of their two-point conversions. The expected value of a two-point conversion attempt is

$$(2 \times .2) + (0 \times .8) = .4.$$

This means that it is better, on average, for Jacksonville to always kick an extra point.

Example: Going for 2

On the other hand, during the 2015 season, the Pittsburgh Steelers made 94.1% of their extra point kicks. Thus, the expected value of an extra point is

$$(1 \times .941) + (0 \times .059) = .941$$

Pittsburgh made 72.7% of their two-point conversions. Thus, the expected value of a two-point conversion attempt is

$$(2 \times .727) + (0 \times .273) = 1.454$$

This means that, on average, it is better for Pittsburgh to go for 2 on every attempt.

Risk and Utility

Today's Goals

- We will study how mathematicians measure risk.
- We will apply the idea of expected value to studying utility and discuss the St. Petersburg paradox.

Expected value

Expected value is the probabilistic version of a weighted average.

Suppose that X is a variable that takes on the numerical values (outcomes) $x_1, x_2, \ldots, x_N$ with probabilities $p_1, p_2, \ldots, p_N$, respectively.

The **expected value** (or expectation) for X is given by

$$E = x_1 \cdot p_1 + \cdots + x_N \cdot p_N.$$

That is, it is the sum of each outcome times its probability.

Risk: Stock Market

Say a pharmaceutical company's stock is currently selling for \$10. It has a new drug that must be approved by the FDA before it can be sold to the general public. There are three phases to the trials (Phase I, Phase II, Phase III).

If the drug passes all the phases and the drug goes to the market, the stock price is projected to jump to \$100.

On the other hand, if it fails Phase I then the drug is considered a flop and the price plummets to \$0.

If the drug passes Phase I but fails Phase II then the price will drop to \$5. If the drug passes Phase II but fails Phase III there is hope for the drug in the future so the price rises to \$15.

Risk: Stock Market

Based on historical averages:
The probability that the drug fails Phase I is 25%.
The probability that it passes Phase I but fails Phase II is 45%.
The probability that it passes Phase II but fails Phase II is 20%.
The probability that it passes Phase III is 10%.

What is the expected value of the company's stock?

$$E = 0.25 \times \$0 + 0.45 \times \$5 + 0.20 \times \$15 + 0.10 \times \$100 = 15.25.$$

Risk: Stock Market

Is it worth purchasing the stock, given that the current price is \$10 and the expected future value is \$15? The answer depends on a few variables.

It may be worth purchasing the stock so long as one's portfolio is diverse enough to absorb the loss should the drug fail in Phase I or II.

It may not be worth purchasing the stock given that it will take up to several years to realize any gains even if the drug passes Phase I and Phase II. One could make similar or even better gains (then the expected value) in a more stable investment.

Risk: Setting life insurance premiums

A life insurance company is offering a \$100,000 one-year term life insurance policy to a 55-year-old nonsmoking female in moderately good health. What should be a reasonable premium for this policy?

Let \$P be the break-even (or fair) premium that the insurance company should charge. In other words, \$P is what the insurance company should charge so that the expected value of the insurance is \$0.

Risk: Setting life insurance premiums

Suppose that based on historical averages, the probability that a 55-year-old nonsmoking female in moderately good health will die is 0.2%. If she were to die, the payoff would be \$(P-100,000). If she doesn't die, the payoff would be \$P. Thus, we want to solve the equation

$$0 = 0.002 \times \$(P - 100,000) + 0.998 \times \$P.$$

This works out to $P = 200$.

This means that the insurance company should charge \$200 for the policy if it wants to break even and more if it wants to generate a profit. How much more depends on the market.

Risk: Coin flipping

You are invited to play a game. In this game, you are given \$50 that you may keep or gamble on a coin flip. However, you have the option to flip a coin. If it lands tails, you lose your \$50.

The expected value of this game is $E = \frac{1}{2} \cdot \$100 + \frac{1}{2} \cdot \$0 = \$50$, exactly the same as keeping the \$50. What if the guaranteed money was more or less than \$50?

How you would play this game depends upon your *risk attitude*.

- risk-averse: accept a certain payment of less than \$50.
- risk-neutral: indifferent between the bet and certain payment.
- risk-loving: accept the bet even when the guaranteed payment is more than \$50.

The St. Petersburg Paradox

Suppose you play a game with a series of coin flips. You get to flip a coin until it comes up heads.

If heads comes up on the first flip (1/2 chance), you win \$1.
If heads comes up on the second flip (1/4 chance), you win \$2.
If heads comes up on the third flip (1/8 chance), you win \$4.
And so on, if heads comes up on the nth flip, you win $\$2^{n-1}$.

What is the right entrance fee (initial bet) that one should be willing to pay to play this game?

The St. Petersburg Paradox

We will compute the expected value of this game.

$$
\begin{aligned}
E &= \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 4 + \cdots \\
&= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots
\end{aligned}
$$

This is an infinite series that is *divergent* (the sum is ∞).

This means that one should be willing to pay any amount to play the game.

That cannot possibly be right.

The St. Petersburg Paradox

This problem was devised by Nicolas Bernoulli and a solution was presented by his cousin, Daniel Bernoulli, thirty years later.

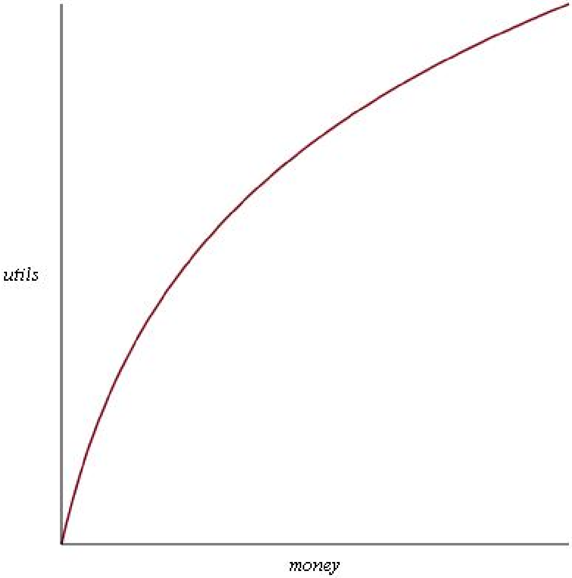
The solution proposed argues that a dollar is not a dollar is not a dollar (technically he worked in *ducats*, a commonly accepted currency in Europe at the time).

As Ellenberg points out, winning \$2000 is not twice as good as winning \$1000, it's less than twice as good because \$1000 is worth less to someone who already has \$1000 than it is to someone who has \$0.

The St. Petersburg Paradox

Daniel Bernoulli argued using *utility*, which is measured in terms of *utils*. He claimed that the relation between utility and money grew like a logarithm.

He claimed $\$2^k$ is worth k utils.



The St. Petersburg Paradox

Thus, according to Daniel Bernoull, the expected value (in terms of utils) is

$$
\begin{aligned}
E &= \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \cdots \\
&= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \cdots \\
&= 2
\end{aligned}
$$

According to our relationship before, 2 utils is worth $\$2^2 = \$4$.

Birthday Problem

Some announcements

- Reading reflection for Chapter 13 due Monday, March 13
- Announcements for game due (via email) on Wednesday, March 15
- Homework 6 due on March 15
- Exam 3 on March 17

Today's Goals

- The Birthday Game (fun with probability).

50 Cent

This is 50 Cent.

He's going to party like it's your birthday.

The Birthday Game

What is the probability that two people in the class have the same birthday?

We will go by date, ignoring year and also ignoring leap day (sorry leap babies).

The Birthday Game

It turns out it is easier to compute the complementary event: What is the probability that no two people have the same birthday?

Say we start with just two people. What is the probability that their birthdays are on different days?

$$P = \frac{365}{365} \times \frac{364}{365} \approx 99.73\%$$

The Birthday Game

What is the probability that three people have their birthdays, all on different days?

$$P = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \approx 99.18\%$$

That's still pretty likely. What if we jump to 10 people?

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{356}{365} \approx 88.31\%$$

The Birthday Game

15 people?

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{351}{365} \approx 74.71\%$$

At this point there is about a 75% chance that two people *do not* share the same birthday.

18 people?

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{348}{365} \approx 65.31\%$$

So if everyone is here today, there's about a 65% chance two people *do not* share a birthday, so about a 35% chance that two people do share a birthday.

The Birthday Game

21 people?

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{345}{365} \approx 55.63\%$$

22 people?

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{344}{365} \approx 52.43\%$$

23 people?

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{343}{365} \approx 49.27\%$$

This means that in a group of 30 people, it is more likely than not that two people will share the same birthday.

The Birthday Game

This game works best in a class of about 30 people.

$$P = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{336}{365} \approx 29.37\%$$

In this case, there's over a 70% chance that two people share a birthday.

Hamming Codes

Some announcements

- Announcements for game due (via email) on Wednesday, March 15
- Homework 6 due on March 15
- Exam 3 on March 17

Today's Goals

- Learn about error correcting codes and how they work.
- Construct Hamming Codes and use them to transmit simple messages.
- Examine how these codes are relevant to optimal strategies for selecting lottery tickets.

Hamming Codes

Hamming codes are a way to transmit messages that may contain errors and correct for them.

Note: this is different then transmitting messages secretly (RSA), though the two may be used together.

Hamming codes are written in *bits*. A **bit** is a digit which is either zero or one. This is also called *binary* or *mod 2*.

A **string** is a sequence of bits.

Attached to each string in a Hamming code is an additional parity check string (also in bits).

Hamming Codes Example

We begin with a 4-bit string. Call the string a1a2a3a4. Three check digits, c1, c2, and c3, will be attached to the 4-bit string to produce a 7-bit string.

c1, c2, and c3 are chosen as follows:

$$c1 = \begin{cases} 0 & \text{if } a1 + a2 + a3 \text{ is even} \\ 1 & \text{if } a1 + a2 + a3 \text{ is odd} \end{cases}$$

$$c2 = \begin{cases} 0 & \text{if } a1 + a3 + a4 \text{ is even} \\ 1 & \text{if } a1 + a3 + a4 \text{ is odd} \end{cases}$$

$$c3 = \begin{cases} 0 & \text{if } a2 + a3 + a4 \text{ is even} \\ 1 & \text{if } a2 + a3 + a4 \text{ is odd} \end{cases}$$

Hamming Codes-Construction

The sums shown above are called parity check sums - so named because their purpose is to ensure that the sum of various components of the encoded message is even.

For example, c1 is defined in so that $a1 + a2 + a3 + c1$ is even

Example

Construct the Hamming code word corresponding to the 4 bit string 0101

$$a1 + a2 + a3 = 0 + 1 + 0 \text{ is odd so } c1 = 1$$
$$a1 + a3 + a4 = 0 + 0 + 1 \text{ is odd so } c2 = 1$$
$$a2 + a3 + a4 = 1 + 0 + 1 \text{ is even so } c3 = 0.$$

The Hamming code word corresponding to 4-bit string 0101 is 0101110.
Every possible sequence of 7 bits is either a correct message (corresponds to a Hamming code word) or contains exactly one correctable error.

Error Detection and Error Correction

Sometimes, due to noisy transmission, code words contain errors. The Hamming Code is designed to detect and correct errors in 4-bit transmissions.

Suppose a message is received as 1111010. Is this a Hamming code word? If not, what word should it have been?

Error Detection and Error Correction

In order to determine if the message received is a Hamming Code word, we simply scan the code. If it is one of the 16 code words, we know the message is received as sent.

If it is not among the 16 code words, we compare the message received with each code word and compute the Hamming distance for each.

The Hamming distance is defined as the number of times a bit in the received message differs from the bit in the code word.

Example

Compare the code words 0001011 with 1111010.

These words differ in 4 positions.

We say the **Hamming distance** is 4.

Error Detection and Error Correction

Once all the distances are computed, we locate the Hamming code which produces the shortest distance for 1111010. We also call this the **nearest** code word. This code will be the code used to correct the transmission error.

If there is more than one shortest distance, we do not correct the message.

The Transylvania Lottery

This game is like a usual pick 3 from the numbers 1 through 7 (no repeats) with a jackpot for hitting all three and a smaller prize for hitting two out of three (*deuce*).

The Transylvania Lottery

How many possible combinations are there of tickets?

$$_7C_3 = \frac{7!}{3!(7-3)!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2} = 35.$$

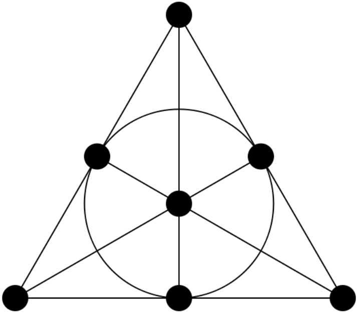
Say you are going to buy seven tickets. What are your chances of hitting the Jackpot?

$$\frac{7}{35} = \frac{1}{5} = 20\%.$$

This is true no matter what tickets you choose.

Say you wanted to maximize the number of deuces? Does it matter what tickets you select? Yes!

The Fano Plane

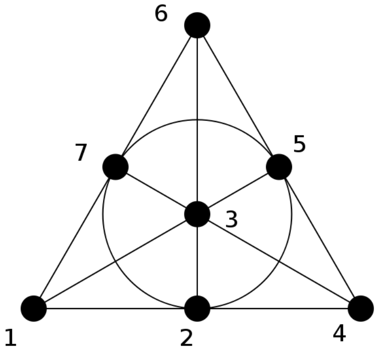


This is the Fano plane. It is an example of a finite geometry (seven points). It satisfies two rules:

(1) every pair of points is contained in exactly one common line and

(2) every pair of lines contains exactly one common point.

The Fano Plane



There are exactly 7 lines in the Fano plane. They are:

124, 135, 176, 236, 257, 437, 456

These are *exactly* the numbers we should pick for the Translyania Lottery.

The Transylvania Lottery

Why these numbers?

$$124, 135, 176, 236, 257, 437, 456.$$

Try picking any three numbers.

There are either three deuces or no deuces.

But, if there are no deuces then we win the jackpot!

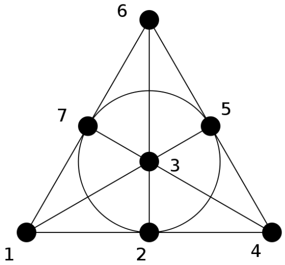
Hamming Codes

What does this have to do with Hamming codes?
Here's another codebook from the Ellenberg text:

$000 \to 0000000$	$001 \to 0010111$
$010 \to 0101011$	$011 \to 0111100$
$101 \to 1011010$	$110 \to 1100110$
$100 \to 1001101$	$111 \to 1110001$

This one satisfies a special property that no code with one error is close to two different codes from the codebook so we can always figure out the correct message from Hamming distance.

Hamming Codes and the Fano Plane



Say we translate each line in the Fano Plane to binary. We put a 0 in the n spot if point n is on the line and a 1 if it is not.

Then the line 124 is 0010111.

The line 257 is 1011010.

You can check that each line corresponds to one code from the codebook.

Linear Algebra

Some announcements

- Game Day Reports (1 per group) due on 3/31 (printed and handed in)
- Individual reflections on Game Day due 3/31 (Sakai dropbox)

Today's Goals

- As a prerequisite to discussing linear regressions more formally, we will learn some tools of linear algebra.
- Learn basic algebraic operations on vectors and matrices.
- Learn how to check whether two vectors are *orthogonal*.

Linear Algebra

Linear algebra is a branch of algebra derived from the study of systems of linear equations. You may remember doing this in high school algebra.

Example. Find the solution(s) to the following system of equations.

$$\begin{aligned} x + y &= 2 \\ x - y &= 6. \end{aligned}$$

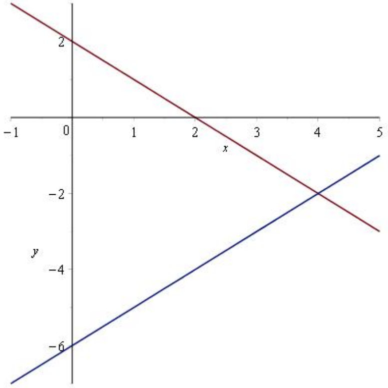
Adding the two equations together gives $2x = 8$, so $x = 4$. Now substituting 4 in for x in either equation gives $y = -2$.

Linear Algebra

Geometric viewpoint:

The equation $x + y = 2$ represents the line $y = -x + 2$ with slope -1 and y-intercept 2.

The equation $x - y = 6$ represents the line $y = x - 6$ with slope 1 and y-intercept -6. Notice where the lines intersect.



Linear Algebra

Linear algebra may have been derived from the study of such problems, but its applications are far reaching.

For example, Google's search algorithm uses a linear algebra tool known as *eigenvectors* to rank pages in searches.

We won't get to discussing eigenvectors but we will study vectors and matrices so that we can use the language and tools of linear algebra to construct linear regressions.

Vectors

Trying to visualize four or five dimensions is difficult, if not impossible. However, working in higher dimensions mathematically is often relatively straight-forward.

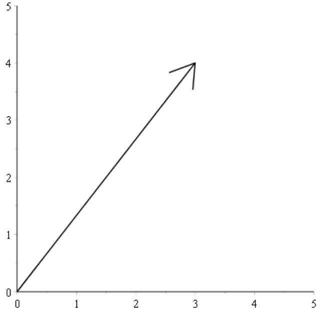
An *n*-dimensional **vector** is a list of *n* numbers.

Examples.
$\langle 1, 5, -2 \rangle$ is a 3-dimensional vector
$\langle 0, 2, 1, 1.29 \rangle$ is a 4-dimensional vector

Vectors

Vectors represents objects in n-dimensional space (also called vectors). These are arrows (or rays) originating at the origin and directed at the point indicated by the entries.

Example.
The 2-dimensional vector $\langle 3, 4 \rangle$ represents the geometric object displayed.



Basic operations on vectors

Vector addition. The sum of two n-dimensional vectors $\vec{u}$ and $\vec{v}$ is the n-dimensional vector $\vec{u} + \vec{v}$ whose kth entry is the sum of the kth entries of $\vec{u}$ and $\vec{v}$.

Example. Add the vectors $\vec{u} = \langle 1, 2, -1 \rangle$ and $\vec{v} = \langle 0, 1, 2 \rangle$.

$$\begin{aligned} \vec{u} + \vec{v} &= \langle 1, 2, -1 \rangle + \langle 0, 1, 2 \rangle \\ &= \langle 1 + 0, 2 + 1, -1 + 2 \rangle \\ &= \langle 1, 3, 1 \rangle. \end{aligned}$$

Geometrically, this is the *parallelogram rule*.

Basic operations on vectors

Scalar multiplication. The product of an n-dimensional vector $\vec{v}$ and a real number c is the n-dimensional vector $c\vec{v}$ whose kth entry is the product of c and the kth entry of $\vec{v}$.

Example. Multiply $\vec{v} = \langle 1, 2, 0 \rangle$ and $c = 2$.

$$c\vec{v} = 2\langle 1, 2, 0 \rangle = \langle 2 \cdot 1, 2 \cdot 2, 2 \cdot 0 \rangle = \langle 2, 4, 0 \rangle.$$

Geometrically we think of this as stretching or contracting the vector $\vec{v}$ by a factor of c.

Vector algebra

We can put these two rules together as well.
Example. Let $\vec{u} = \langle 1, 5, -2 \rangle$, $\vec{v} = \langle 2, -1, 3 \rangle$, and $c = 2$. Compute $c(\vec{u} + \vec{v})$ and $c\vec{u} + c\vec{v}$.

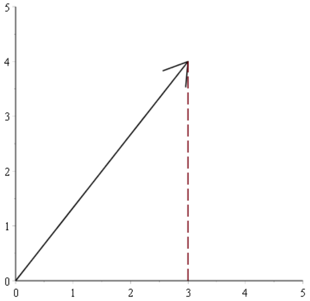
$$
\begin{aligned}
&c(\vec{u} + \vec{v}) \\
&\quad = 2\left(\langle 1, 5, -2 \rangle + \langle 2, -1, 3 \rangle\right) \\
&\quad = 2\langle 3, 4, 1 \rangle \\
&\quad = \langle 6, 8, 2 \rangle.
\end{aligned}
$$

$$
\begin{aligned}
&c\vec{u} + c\vec{v} \\
&\quad = 2\langle 1, 5, -2 \rangle + 2\langle 2, -1, 3 \rangle \\
&\quad = \langle 2, 10, -4 \rangle + \langle 4, -2, 6 \rangle \\
&\quad = \langle 6, 8, 2 \rangle.
\end{aligned}
$$

Note: The associative, commutative, and distributive laws all hold for these operations.

Length of vectors

The **length** of a vector is the length of the corresponding geometric object. To compute this, we can use the Pythagorean Theorem.



Example. $\vec{v} = \langle 3, 4 \rangle$.

The 2-dimensional vector $\vec{v}$ is the hypotenuse of a right triangle with base 3 and height 4.

By the Pythagorean Theorem, the length of $\vec{v}$ is 5.

Length of vectors

The Pythagorean Theorem generalizes to any dimension and so we can define length in the following way.

The **length** of an *n*-dimensional vector $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ is

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Example. Find the length of the vector $\vec{v} = \langle -1, 0, 4 \rangle$.

$$\|\vec{v}\| = \sqrt{(-1)^2 + (0)^2 + (4)^2} = \sqrt{1 + 0 + 16} = \sqrt{17} \approx 4.123.$$

Distance

With this formula in hand, we can now define distance. This generalizes the formula for distance between points.

The **distance** between vector $\vec{u}$ and $\vec{v}$, written $\mathrm{dist}(\vec{u}, \vec{v})$, is the length of the vector $\vec{u} - \vec{v}$. That is,

$$\mathrm{dist}(\vec{u}, \vec{v}) = \left\| \vec{u} - \vec{v} \right\| .$$

Example. Find the distance between $\vec{u} = \langle 1, 5, 0 \rangle$ and $\vec{v} = \langle -1, 4, 2 \rangle$.

$$\mathrm{dist}(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\| = \|\langle 2, 1, -2 \rangle\| = \sqrt{4 + 1 + 4} = 3.$$

Orthogonality and the dot product

In two dimensions, we say two lines are orthogonal if they are perpendicular. That is, they intersect at a right angle. Thus, to determine if two vectors are orthogonal, we need to be able to measure the angle between them. The dot product is one tool for this.

The **dot product** of two n-dimensional vectors $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ is

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Note: The dot product is a number, not a vector. Also, it plays nice with the operations of addition and scalar multiplication.

Orthogonality and the dot product

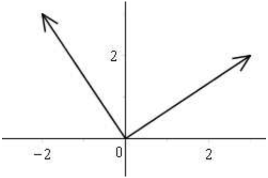
Example. Find the dot product of $\vec{u} = \langle 1, 2, -2 \rangle$ and $\vec{v} = \langle 3, 0, 1 \rangle$.

$$\vec{u} \cdot \vec{v} = (1 \cdot 3) + (2 \cdot 0) + (-2 \cdot 1) = 3 + 0 - 2 = 1.$$

Example. Find the dot product of $\vec{u} = \langle 3, 2 \rangle$ and $\vec{v} = \langle -2, 3 \rangle$.

$$\vec{u} \cdot \vec{v} = (3 \cdot -2) + (2 \cdot 3) = -6 + 6 = 0.$$

Note that these vectors intersect in a right angle.



Orthogonality and the dot product

Two n-dimensional vectors $\vec{u}$ and $\vec{v}$ are **orthogonal** if $\vec{u} \cdot \vec{v} = 0$.

Linear Algebra
Part Deux

Some deadlines

- Monday: nothing due, classwork practice on linear algebra
- Tuesday: reflections due, reading discussion
- Wednesday: HW due, Maple lab (must have Maple installed on your laptop by this day and bring laptops to class)
- Friday: Game Day Reports (1 per group, printed) and Individual reflections (Sakai dropbox) due, Linear regression lecture

Today's Goals

- As a prerequisite to discussing linear regressions more formally, we will learn some tools of linear algebra.
- Last time we learned basic operations on vectors. Today we'll study matrices.
- We will learn how to add and multiply matrices, as well as find the transpose and inverse of a matrix.

Matrices

A **matrix** is an array of data, arranged in rows and columns. We say a matrix is of dimension $m \times n$ if it has m rows and n columns.

Examples.

$$\begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 5 \end{bmatrix} \qquad \begin{bmatrix} 6 & 5 \\ 2 & -2 \\ 0 & 0 \end{bmatrix}$$

2×3 matrix $\qquad$ 3×2 matrix

The i, j entry of a matrix is located in the ith row and jth column.

For example, the $1, 2$ entry in 2×3 matrix above is 0 and the $2, 3$ entry is 5.

Matrices

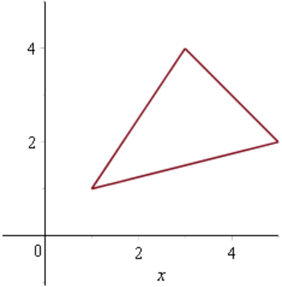
Matrices are a good way to visualize large sets of data. Anyone who has ever worked in a spreadsheet program (like Excel) has used a matrix.

In computer graphics, images are stored as matrices according to their x and y coordinates (x, y, z for 3-dimensional images).

Example. This triangle would be stored as the matrix

$$\begin{bmatrix} 1 & 3 & 5 \\ 1 & 4 & 2 \end{bmatrix}$$

where the rows represent the x and y coordinates (respectively) and the columns represent the different vertices.



Basic operations on matrices

Matrix addition. The sum of two $m \times n$ matrices M and N is the $m \times n$ matrix $M + N$ whose i, j entry is the sum of the i, j entries of M and N.

Example. Add the matrices $M = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix}$ and $N = \begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix}$.

$$M + N = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix} + \begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3+1 & 2+5 \\ 1-2 & 5+3 \\ -1+1 & -3+2 \end{bmatrix} = \begin{bmatrix} 4 & 7 \\ -1 & 8 \\ 0 & -1 \end{bmatrix}.$$

Note: If the two matrices are not the same size then their sum is not defined.

Basic operations on matrices

Scalar multiplication. The product of an $m \times n$ matrix M and a real number c is the $m \times n$ matrix cM whose i, j entry is the product of c and the i, j entry of M.

Example. Multiply $M = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix}$ and $c = 2$.

$$cM = 2 \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 & 2 \cdot 2 \\ 2 \cdot 1 & 2 \cdot 5 \\ 2 \cdot (-1) & 2 \cdot (-3) \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 2 & 10 \\ -2 & -6 \end{bmatrix}.$$

Note: We can combine operations as well. The associative, commutative, and distributive laws all hold for these operations.

Matrix multiplication

Recall that we discussed how to multiply vectors using the dot product and that the output was a scalar. We will use this operation to obtain the product of two matrices under certain conditions.

Matrix multiplication. The product of an $m \times k$ matrix M and a $k \times n$ matrix N is the $m \times n$ matrix MN whose i, j entry is the dot product of row i of M and column j of N.

Note: One requirement of this definition is that the number of *columns* of M is the same as the number of *rows* of N.

Matrix multiplication

Example. Let $M = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$, $N = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix}$. Find MN and NM.

$$MN = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 0 & 1\cdot 1 + 2\cdot 3 \\ 1\cdot 1 + 0\cdot 0 & 1\cdot 1 + 0\cdot 3 \end{bmatrix} = \begin{bmatrix} 1 & 7 \\ 1 & 1 \end{bmatrix}$$

$$NM = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 1\cdot 1 & 1\cdot 2 + 1\cdot 0 \\ 0\cdot 1 + 3\cdot 1 & 0\cdot 2 + 3\cdot 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 3 & 0 \end{bmatrix}$$

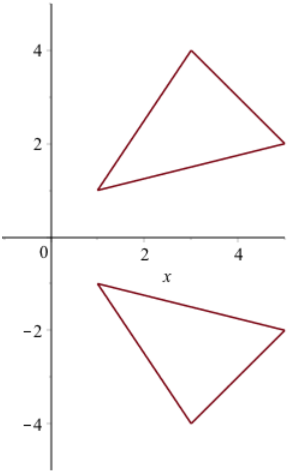
We say matrix multiplication is *not commutative* because $MN \neq NM$ in general.

Computer graphics redux

Say we wanted to tell a computer to reflect our triangle across the x-axis. One way to do this would be to instruct it to move each point individually, or we could give it a rule that tells the computer how it should transform the x-coordinates and how it should transform the y-coordinates.

This transformation could be represented as

$$T(x, y) = (x, -y).$$



Computer graphics redux

Recall that our triangle is represented as $\begin{bmatrix} 1 & 3 & 5 \\ 1 & 4 & 2 \end{bmatrix}$.

We can represent the transformation $T(x, y) = (x, -y)$ as the matrix $S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Note that

$$SM = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 \\ 1 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 5 \\ -1 & -4 & -2 \end{bmatrix}.$$

These are exactly the coordinates of the reflected triangle.

What is MS?

Transpose of a matrix

Matrix transpose. The transpose of an $m \times n$ matrix M is an $n \times m$ matrix M^T whose ith row is the ith column of M.

Example. Find the transpose of $M = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 4 & 2 \end{bmatrix}$.

$$M^T = \begin{bmatrix} 1 & 1 \\ 3 & 4 \\ 5 & 2 \end{bmatrix}$$

Note that $(M^T)^T = M$.

The identity matrix

The matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the (2×2) **identity matrix**. It is so-named because $IM = M$ and $MI = M$ for *any* 2×2 matrix M.

Question. Given a 2×2 matrix M, can we find another matrix N such that $MN = I$?

Answer. Sometimes.

A 2×2 matrix $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is **invertible** if $ad - bc \neq 0$.

Inverse of a 2×2 matrix

Matrix inverse. The inverse of an 2×2 matrix $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$M^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Example. Find the inverse of $M = \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}$.

Since $ad - bc = 1 \cdot 4 - 1 \cdot 2 = 2 \neq 0$, then M is invertible and

$$M^{-1} = \frac{1}{2} \begin{bmatrix} 4 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1/2 \\ -1 & 1/2 \end{bmatrix}.$$

We can check this is right by performing the matrix multiplication MM^{-1} and checking that it is the identity matrix.

Linear Regression

Some announcements/deadlines

I posted some additional matrix multiplication practice on Sakai. This is not required but is there if you need it. You can check your answers in Maple.

Next week:

- Monday: second Maple lab, bring laptops
- Tuesday: reflections due, reading discussion
- Wednesday: classwork
- Friday: Exam 4, no calculators!

Today's Goals

- We will discuss systems of equations that have no solutions (inconsistent systems).
- We will learn how to approximate solutions to inconsistent systems using linear algebra.
- We will apply this to finding lines of linear regression.

Systems of equations

Recall from the Maple lab how matrices and vectors can be used to solve a system of (linear) equations.

Example. The system

$$\begin{aligned} 3x + y &= 5 \\ 2x + y &= -1 \end{aligned}$$

can be written as

$$\begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}.$$

To solve the system, we multiply both sides by the inverse of the 2×2 matrix.

Systems of equations

Example (cont).

$$\begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ -13 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ -13 \end{bmatrix}.$$

Thus, the solution to the system is $x = 6$, $y = -13$.

Systems of equations

This example illustrates a general method for solving systems of equations. Let A be the **coefficient matrix** of the system, $\mathbf{x}$ the vector of indeterminates/variables and $\mathbf{b}$ the **constant vector**.

Example. For the system

$$\begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix},$$

we have

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}.$$

Systems of equations

Given the **matrix equation** $A\mathbf{x} = \mathbf{b}$ where A is $n \times n$ and both $\mathbf{x}$ and $\mathbf{b}$ are n-dimensional vectors, we can solve by multiplying both sides by A^{-1}.

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^{-1}A\mathbf{x} &= A^{-1}\mathbf{b} \\ I\mathbf{x} &= A^{-1}\mathbf{b} \\ \mathbf{x} &= A^{-1}\mathbf{b}. \end{aligned}$$

Question. Does this always work?

Answer. No. Not all matrices have an inverse.

Systems of equations

Consider the system of equations

$$\begin{aligned} x + y &= 1 \\ x + y &= 3. \end{aligned}$$

This system has no solutions. We say it is **inconsistent**. Geometrically we can think of this as a pair of parallel lines.

However, one *could* ask what points come closest to being solutions to the system.

Least-squares solution

A **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an n-dimensional vector $\hat{\mathbf{x}}$ such that

$$\|\mathbf{b} - A\hat{x}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

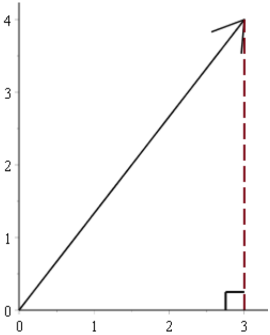
for *all* n-dimensional vectors $\mathbf{x}$.

Geometrically, we are thinking of $A\mathbf{x}$ as a space of vectors and $\mathbf{b}$ as a point not in that space. To find the least squares solution means to find the point in that space that is closest to $\mathbf{b}$. In order to do this, we need to *project*.

Projections

Projections are relatively simple in two-dimensions.

Example. What is the point on the x-axis closest to the point $(3, 4)$? It's clear that the point is $x = 3$. This corresponds to the points of intersection found by drawing a line from $(3, 4)$ that is *perpendicular (orthogonal)* to the x-axis.



Least-squares solution

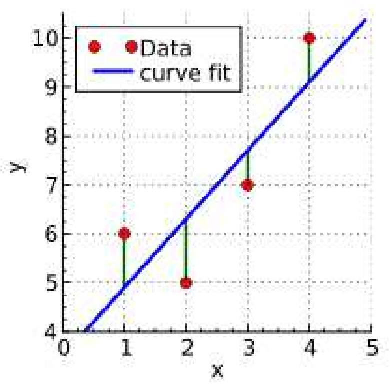
The **least-squares solution** $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ is given by

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

The distance $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the **least-squares error**.

Line of regression

Given a set of data points: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we want to the point that best fits this set of data. First we need to say what that means.



Given a line that we think best fits the data, define the **residual** to be the length of the vertical line from a point to that line.

The **line of regression** is the line that *minimizes* the sum of the squares of the residual.

Line of regression

Given a set of data points: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we will find the **line of regression**, also called the **least-squares line**

$$y = \beta_0 + \beta_1 x.$$

Here, β_1 is the slope and β_0 is the y-intercept.

Suppose the given points were all on a line together. Then obviously $y = \beta_0 + \beta_1 x$ would be that line and so we would have

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 \\ y_2 &= \beta_0 + \beta_1 x_2 \\ \vdots \quad &\quad \vdots \\ y_n &= \beta_0 + \beta_1 x_n. \end{aligned}$$

Line of regression

We could rewrite the system of equations from the previous slide as

$$X\boldsymbol{\beta} = \mathbf{y}$$

where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

We refer to X as the **design matrix**, $\boldsymbol{\beta}$ as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

Line of regression

If the points *are not* on a line together (as we would expect), then the system $X\beta = \mathbf{y}$ is inconsistent. Said another way, there is *no* β that makes this statement true.

However, computing the least-squares solution of $X\beta = \mathbf{y}$ is equivalent to finding the β that determines the line of regression (least-squares line).

In summary, the y-intercept (β_0) and the slope (β_1) of the line of regression are given by

$$\beta = (X^T X)^{-1} X^T \mathbf{y}.$$

Line of regression

Example. Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(4,1)$, $(1,2)$, $(3,3)$, $(5,5)$.

We build the matrix X and vector $\mathbf{y}$ from the data,

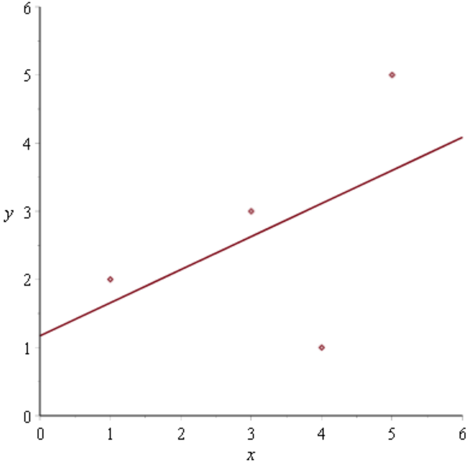
$$X = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}.$$

Using Maple we find that

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (XX^T)^{-1}X^T\mathbf{y} = \begin{bmatrix} 41/35 \\ 17/35 \end{bmatrix}.$$

Line of regression

Thus, the line of regression is $y = \frac{41}{35} + \frac{17}{35}x$.



Cause and Effect

Some announcements/deadlines

Remember that we have Exam 4 on Friday. No calculators!

Today's Goals

- Discuss correlation in more detail.
- Consider some examples of correlation vs causation.
- Discuss Berkson's fallacy.

Some examples in today's lecture are borrowed from *Excursions in Mondern Mathematics*. The text has some other good examples as well.

Hormone Replacement Therapy

Hormone replacement therapy (HRT) is a form of estrogen replacement used to treat symptoms of menopause in women. The use of this therapy is not considered very controversial.

Question: Should women continue HRT after menopause?

There have been a number of conflicting answers to this question.

Hormone Replacement Therapy

Nurses Health Study (1985): Women taking estrogen had one-third the number of heart attacks than women who didn't.

HRT becomes a highly recommended treatment for preventing heart disease and osteoporosis in post-menopausal women.

By 2001, 15 million prescriptions for HRT are filled annually with about 5 million for post-menopausal women.

Hormone Replacement Therapy

Women's Health Initiative Study (2002): HRT significantly increases the risk of heart disease and breast cancer, and is a risk factor for stroke.

The only health benefit of HRT is for preventing osteoporosis and possibly colorectal cancer.

HRT is no longer recommended as a general therapy for older women but still recommended as an effective for women during menopause.

Hormone Replacement Therapy

Women's Health Initiative Follow-up (2007): HRT offers protection against heart disease for women that start taking it during menopause but increases the risk of heart disease for women who start taking it after menopause.

Women's Health Initiative Follow-up II (2009): HRT significantly increase the risk of breast cancer in menopausal and post-menopausal women.

The story of HRT demonstrates the difficulty in determining cause-and-effect. In other words, determining *correlation* is easy. Determining *causation* is hard.

Correlation vs Causation

A **correlation** between two events X and Y occurs when there is an observed mutual relationship between the two events.

In terms of a scatterplot, this means that there is an ellipse with **high eccentricity** (skinny ellipse) that encloses the data as opposed to an ellipse with **low eccentricity** (fat ellipse, or a circle).

A **positive correlation** (ellipse oriented SE to NW) means that a rise in X implies a rise in Y. A **negative correlation** (ellipse oriented NE to SW) means that a rise in X implies a decrease in Y.

A **causation** between X and Y occurs when one event is the cause of the other.

Correlation vs Causation

Once a correlation is established, $X - Y$, there are several possible explanations.

- X is the cause of Y: $X \Rightarrow Y$
- Y is the cause of X: $Y \Rightarrow X$
- X and Y have a common cause, C:

$$\begin{tikzcd} & C \arrow[dl, Rightarrow] \arrow[dr, Rightarrow] & \\ X \arrow[rr, dash] & & Y \end{tikzcd}$$

- X and Y have separate causes, C_1, C_2:

$$\begin{tikzcd} C_1 \arrow[r, dash] \arrow[d, Rightarrow] & C_2 \arrow[d, Rightarrow] \\ X \arrow[r, dash] & Y \end{tikzcd}$$

Does drinking coffee help you live longer?

NIH-AARP Diet and Health Study (2012): This was a large *observational study* involving 400,000 subjects living in the US aged 50 to 71. Subjects were given an extensive survey concerning their health, nutrition, and lifestyle habits. They were then tracked for 12 years for health status, life span, and causes of death if they had died.

Does drinking coffee help you live longer?

One of the lifestyle factors considered was coffee consumption. The initial correlation found by researchers was that the non-coffee drinkers lived longer than coffee drinkers.

During the 12-year period of the study:

- 19% of the male and 15% of the female coffee-drinkers died.
- 13% of the male and 10% of the female non-coffee drinkers died.

These correlations point toward the hypothesis that coffee is bad for you.

Does drinking coffee help you live longer?

Except, coffee drinking was also found to be positively correlated with smoking (coffee drinkers are more likely to be smokers than non-coffee drinkers).

Similar correlations were found between coffee drinking and other lifestyle factors such as consuming more alcohol, consuming more red meat, having lower levels of physical activity, and consuming fewer fruits and vegetables.

Does drinking coffee help you live longer?

Once the researchers *controlled* for smoking (comparing longevity among non-smoking coffee drinkers with longevity among non-smoking non-coffee drinkers) coffee consumption was correlated with living longer and with a reduced incidence of cancer and heart disease.

The same positive correlation showed up when researchers controlled for other *confounding* variables: alcohol consumption, body mass index, age, ethnicity, marital status, physical activity, consumption of red meat, and (for post-menopausal women) use or nonuse of HRT.

Does drinking coffee help you live longer?

The study's conclusion was that men who drank coffee had a 10% better chance of living through the study than those who didn't. Women coffee drinkers had a 13% better chance.

However, this study is not without its flaws and its results are not conclusive. Additional studies will look to clarify some of these relationships.

Correlations from a common cause

$$\begin{tikzcd} & C \arrow[dl, Rightarrow] \arrow[dr, Rightarrow] & \\ X \arrow[rr, dash] & & Y \end{tikzcd}$$

Example. (from Wikipedia): Sleeping with one's shoes on is strongly correlated with waking up with a headache. Therefore, sleeping with one's shoes on causes headache.

A more plausible explanation is that both are caused by a third factor, in this case going to bed drunk, which thereby gives rise to a correlation. So the conclusion is false.

Correlations from a common cause

Example. (from *HNTBW*) A town has a population of 1,000 people. Of these,

- 300 have high blood pressure,
- 400 have diabetes, and
- 120 have both.

Suppose all sick people wind up in the hospital. Then in our hospital population we have

- 180 with high blood pressure but not diabetes,
- 280 have diabetes but not high blood pressure, and
- 120 have both.

Of the 180 nondiabetics in the hospital, 100% have high blood pressure. One might conclude from this that high blood pressure prevents against diabetes.

Correlations from a common cause

This is an example of *Berkson's fallacy*, essentially ignoring that there is a common cause to explain the correlation.

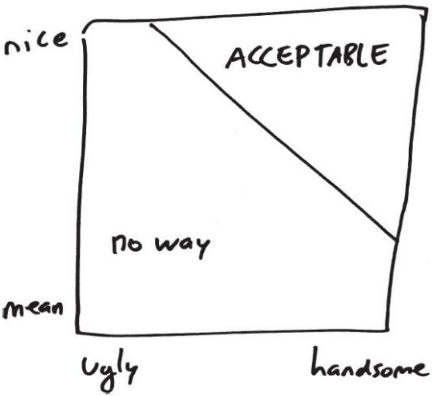
In this case, the common cause is that the two conditions land a person in the hospital.

The Great Square of Men

We assume that men are equally distributed across the square.

Our subject, Alex, only dates men whose niceness plus handomness lies above a certain threshold.

The Great Square of Men

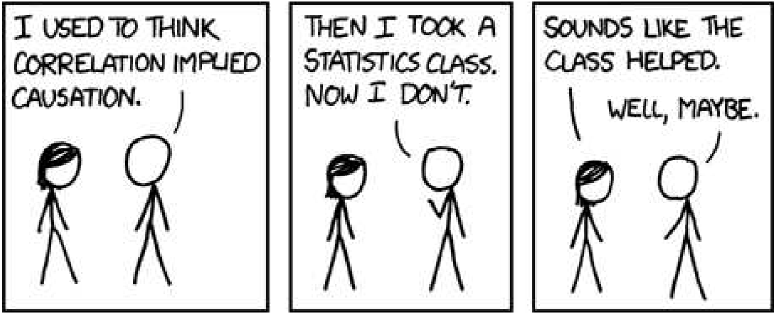


This leads to a smaller Triangle of Acceptable Men.

Alex then might assume that there is a negative correlation between niceness and handsomeness.

However, this is another example of Berkson's fallacy because Alex is only considering acceptable men, and not all men.

Correlation vs Causation



Voting Methods I

Some announcements

- Chapter 17 reflections due tomorrow.
- Project topics due on Wednesday (by email, please).

Today's Goals

- To examine different ways that voters can choose a candidate from a list of options.
- We will discuss advantages and disadvantages to each method.
- Discuss examples from various types of elections.

Terminology

- Candidates
- Voters
- The ballots
- The outcome
- The voting method

Our interest will be in discussing primarily the different types of ballots and outcomes that an election can have. We will only briefly dwell on the other points.

Types of Ballots: Single-Choice Ballot

Choose your favorite flavor of ice cream amongst the options below.

- ○ Chocolate
- ✓ Cookie Dough
- ○ Mint Chocolate Chip
- ○ Vanilla

The winner can be decided by the candidate with the most votes, or via a traditional runoff.

Types of Ballots: Preference Ballot (1)

List the choices of ice cream in order of preference.

1st Cookie Dough

2nd Vanilla

3rd Chocolate

4th Mint Chocolate Chip

Types of Ballots: Preference Ballot (2)

Rank your preferred flavor of ice cream.

3rd	Chocolate
1st	Cookie Dough
4th	Mint Chocolate Chip
2nd	Vanilla

Types of Ballots: Truncated Preference Ballot

Rank your top 3 preferred flavors of ice cream in order of preference.

1st Cookie Dough

2nd Vanilla

3rd Chocolate

Types of Ballots: Approval Voting

Indicate which ice cream flavors you enjoy eating. You may mark as many as you wish, or none at all.

- ✓ Chocolate
- ✓ Cookie Dough
- ○ Mint Chocolate Chip
- ✓ Vanilla

The winner is the candidate with the most votes.

Outcomes

What are some different outcomes that an election can have?

- Winner-only (e.g., modern presidential elections)
- Partial Ranking (e.g., presidential elections prior to the 12th amendment)
- Full Ranking (sometimes just called Ranking) (e.g. College Basketball rankings)

Creating a preference schedule

A preference schedule is an efficient way to record election results using a preference ballot.

Let's take a hypothetical outcome for our ice cream election. We'll abbreviate the names: Cookie Dough (CD), Vanilla (V), Chocolate (C), Mint Chocolate Chip (M).

Creating a preference schedule

Suppose we have the following number of standard preference ballots.

- 18: CD, C, M, V
- 8: V, C, M, CD
- 20: M, C, CD, V
- 15: V, CD, C, M
- 3: C, M, V, CD

Creating a preference schedule

The columns list the different kinds of ballots with the rows indicating the rankings (1st through 4th).

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	V	M	V	C
2nd	C	C	C	CD	M
3rd	M	M	CD	C	V
4th	V	CD	V	M	CD

Creating a preference schedule (alternate)

Now suppose we have the same results but using the alternate preference ballot where voters list their rankings. The order is alphabetical.

- 18: C(2), CD(1), M(3), V(4)
- 8: C(2), CD(4), M(3), V(1)
- 20: C(2), CD(3), M(1), V(4)
- 15: C(3), CD(2), M(4), V(1)
- 3: C(1), CD(4), M(2), V(3)

Creating a preference schedule (alternate)

The columns list number of ballots. The rows list the the candidates in the presecribed order (in this case alphabetical).

Number of voters	18	8	20	15	3
Chocolate	2	2	2	3	1
Cookie Dough	1	4	3	2	4
Mint Chocolate Chip	3	3	1	4	2
Vanilla	4	1	4	1	3

Voting Methods II

Some announcements

- Project topics due today.
- Start working on your ballots and collecting data.

Today's Goals

- To examine some different methods for evaluating election results, including Plurality, Borda Count, Plurality-with-Elimination (IRV), and Pairwise Comparison.
- Work through some examples.
- Discuss how ties can be resolved.

Last Time

We had a running example last time: favorite ice cream flavor. The choices were Chocolate (C), Cookie Dough (CD), Mint Chocolate Chip (M), and Vanilla (V). Here's the preference schedule.

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	V	M	V	C
2nd	C	C	C	CD	M
3rd	M	M	CD	C	V
4th	V	CD	V	M	CD

The Plurality Method

The **Plurality Method** is the most common method used in US elections today. The winner is the candidate with the most votes, regardless of whether this number represents a majority of the voters.

In our ice cream example, Vanilla is the winner using this method. But Vanilla only has 23 out of the 64 1st place votes (about 36%). Note that Vanilla also has the most 4th place votes (38 out of 64 or about 59%).

Plurality Method

Of all the methods, the Plurality Method is the most susceptible to **insincere voting** (also called tactical voting).

Insincere voting is the practice of voting against one's own interest in order to sway the election.

Think: In 2000, some voters who would have preferred Nader voted for Gore because they were afraid of 'wasting their vote'.

We'll now explore different methods to decide a winner in an election.

Borda Count

The **Borda Count Method** requires a preference ballot.

For each ballot, the candidate ranked last gets 1 points. The candidate ranked second to last gets 2 points, and so on. If there are N candidates, the top candidate gets N points.

There are variations on this. For example, we may not give points to all candidates, or we might weight the points differently. This is a **Modified Borda Count**.

2016 Heisman Award

The Heisman Award is decided using a Borda Count Method.

Player	1st (3pts.)	2nd (2pts.)	3rd (1pt.)	Total Points
Lamar Jackson	526	251	64	2144
DeShaun Watson	269	302	113	1,524
Baker Mayfield	26	72	139	361
Dede Westbrook	7	49	90	209
Jabril Peppers	11	45	85	208

Borda Count Example

Number of voters	18	8	20	15	3
1st	CD	V	M	V	C
2nd	C	C	C	CD	M
3rd	M	M	CD	C	V
4th	V	CD	V	M	CD

Vanilla got 23 1st place votes (with 4 points each), 3 3rd place votes (worth 2 points each), and 38 4th place votes (worth 1 point each). Thus the total for Vanilla is 136.

Similarly, Chocolate's total is 180. Mint Chocolate Chip's total is 156 and Cookie Dough's total is 168.

Chocolate wins! But Chocolate came in last using the **Plurality Method**.

Plurality-With-Elimination

Plurality-With-Elimination is also called **Instant Runoff Voting (IRV)**.

The idea is to eliminate the loser from preference ballot and retally the votes. Only 1st place votes are counted.

In the first round, we tally the 1st place votes. A candidate must have a majority to win. Otherwise we eliminate the candidate with the fewest 1st place votes and tally again. This continues until one candidate has a majority (or there is a tie).

IRV example

In our ice cream election, Chocolate has the fewest 1st place votes so we eliminate it.

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	V	M	V	~~C~~
2nd	~~C~~	~~C~~	~~C~~	CD	M
3rd	M	M	CD	~~C~~	V
4th	V	CD	V	M	CD

The effect of this is that Mint Chocolate Chip now has 3 more 1st place votes.

The tally now is Vanilla (23), Mint Chocolate Chip (23), and Cookie Dough (18). No flavor has a majority.

IRV example

Cookie Dough had the fewest 1st place votes in our new tally, so we now eliminate it :-(

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	~~CD~~	V	M	V	~~C~~
2nd	~~C~~	~~C~~	~~C~~	~~CD~~	M
3rd	M	M	~~CD~~	~~C~~	V
4th	V	~~CD~~	V	M	~~CD~~

The effect of this is that Mint Chocolate Chip now has 18 more 1st place votes.

The tally now is Vanilla (23), Mint Chocolate Chip (41).

Mint Chocolate Chip wins!

The Method of Pairwise Comparisons

With a preference schedule, it is possible to compare any pair of candidates by eliminating all others.

In the **Method of Pairwise Comparisons** we compare each pair of candidates, the winner of each comparison is given a point (1/2 point for ties).

The winner is the candidate with the most points.

Pairwise Comparison Example

Number of voters	18	8	20	15	3
1st	**CD**	V	M	V	**C**
2nd	**C**	**C**	**C**	**CD**	M
3rd	M	M	**CD**	**C**	V
4th	V	**CD**	V	M	**CD**

Cookie Dough (33) beats Chocolate (31)

Pairwise Comparison Example

Number of voters	18	8	20	15	3
1st	CD	V	**M**	V	**C**
2nd	**C**	**C**	**C**	CD	**M**
3rd	**M**	**M**	CD	**C**	V
4th	V	CD	V	**M**	CD

Chocolate (44) beats Mint Chocolate Chip (20)

Pairwise Comparison Example

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	**V**	M	**V**	**C**
2nd	**C**	**C**	**C**	CD	M
3rd	M	M	CD	**C**	**V**
4th	**V**	CD	**V**	M	CD

Chocolate (41) beats Vanilla (23)

Pairwise Comparison Example

Number of voters	18	8	20	15	3
1st	**CD**	V	**M**	V	C
2nd	C	C	C	**CD**	**M**
3rd	**M**	**M**	**CD**	C	V
4th	V	**CD**	V	**M**	**CD**

Cookie Dough (33) beats Mint Chocolate Chip (31)

Pairwise Comparison Example

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	**CD**	**V**	M	**V**	C
2nd	C	C	C	**CD**	M
3rd	M	M	**CD**	C	**V**
4th	**V**	**CD**	**V**	M	**CD**

Cookie Dough (38) beats Vanilla (26)

Pairwise Comparison Example

Number of voters	18	8	20	15	3
1st	CD	**V**	**M**	**V**	C
2nd	C	C	C	CD	**M**
3rd	**M**	**M**	CD	C	**V**
4th	**V**	CD	**V**	**M**	CD

Mint Chocolate Chip (41) beats Vanilla (23)

Pairwise Comparison Example

Our tally is then,

- Chocolate (2)
- Cookie Dough (3)
- Mint Chocolate Chip (1)
- Vanilla (0)

Cookie Dough Wins!

So who really won?

We saw 4 different voting methods (Plurality, Borda Count, IRV, Pairwise Comparison). Each gave a different winner.

Several mathematicians, including Condorcet, Borda, and Arrow, have tried to establish fairness criteria for elections. Unfortunately, every voting method violates some fairness criteria. This is the gist behind Arrow's Impossibility Theorem.

Ties

What if there is a tie? The text gives several different ways that a tie can be broken.

- The result is shared (e.g., with some awards)
- Chance (coin flip or card draw)
- Runoff
- Other criteria from the election
- Deferred to another electing body (a tie in the Electoral College results in the House of Representatives selecting the president)

Additional Methods

There are many additional methods. We will discuss three of them:

- approval voting
- least worst defeat
- ranked pairs

Approval Voting

We have already seen an approval ballot in which voters select all acceptable options.

Candidates are given 1 point each time they are selected on a ballot.

The winner is the candidate with the most points, that is, the candidate that is acceptable to the most voters.

Approval Voting Examples

Suppose we have 3 candidates (A,B,C). There are 8 different ballot types, but realistically there are only 6.

Results:

- 15: A,B
- 20: A,C
- 25: B,C
- 10: A
- 15: B
- 15: C

Final Tally: A(45), B(55), C(60). C wins!

Voting Methods III

Some announcements

- Homework due tomorrow (4/19) and final reflections due on Monday (4/24).
- You should be collecting data for the voting projects.
- I am happy to review ballots, discuss your results, or look over a draft of your report at any time.
- Projects are due on May 3 (the day of the final) but may be turned in any time before that.

Today's Goals

- We will discuss two additional methods for analyzing election results with a preference ballot.
- These methods, along with Pairwise Comparison, are known as Condorcet methods and satisfy a fairness criterion known as the Condorcet Criterion.
- We will consider some examples of each.
- We will discuss various fairness criteria and how the different voting methods violate these criteria.

Last Time

We had a running example last time: favorite ice cream flavor. The choices were Chocolate (C), Cookie Dough (CD), Mint Chocolate Chip (M), and Vanilla (V). Here's the preference schedule.

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	V	M	V	C
2nd	C	C	C	CD	M
3rd	M	M	CD	C	V
4th	V	CD	V	M	CD

Condorcet

The mathematician/social scientist/French revolutionist Marie-Jean-Antoine-Nicolas de Caritat, Marquis de Condorcet, established the following axiom of voting:

If the majority of voters prefer candidate A to candidate B, then candidate B cannot be the people's choice

Our example is a violation of this if we allow Vanilla to win.

Condorcet

We call a candidate a **Condorcet Candidate** if they beat every other candidate in a head-to-head (pairwise) matchup. In our ice cream election, Cookie Dough is a Condorcet Candidate.

A method is known as a **Condorcet Method** if a Condorcet Candidate is guaranteed to win using that method.

The Method of Pairwise Comparisons (that we discussed last time) is a Condorcet Method, why?

The Method of Pairwise Comparisons

With a preference schedule, it is possible to compare any pair of candidates by eliminating all others.

In the **Method of Pairwise Comparisons** we compare each pair of candidates, the winner of each comparison is given a point (1/2 point for ties).

The winner is the candidate with the most points.

Pairwise Comparison Example

Number of voters	18	8	20	15	3
1st	CD	V	M	V	C
2nd	C	C	C	CD	M
3rd	M	M	CD	C	V
4th	V	CD	V	M	CD

Cookie Dough (33) beats Chocolate (31)

Pairwise Comparison Example

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	V	**M**	V	**C**
2nd	**C**	**C**	**C**	CD	**M**
3rd	**M**	**M**	CD	**C**	V
4th	V	CD	V	**M**	CD

Chocolate (44) beats Mint Chocolate Chip (20)

Pairwise Comparison Example

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	**V**	M	**V**	**C**
2nd	**C**	**C**	**C**	CD	M
3rd	M	M	CD	**C**	**V**
4th	**V**	CD	**V**	M	CD

Chocolate (41) beats Vanilla (23)

Pairwise Comparison Example

Number of voters	18	8	20	15	3
1st	**CD**	V	**M**	V	C
2nd	C	C	C	**CD**	**M**
3rd	**M**	**M**	**CD**	C	V
4th	V	**CD**	V	**M**	**CD**

Cookie Dough (33) beats Mint Chocolate Chip (31)

Pairwise Comparison Example

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	**CD**	**V**	M	**V**	C
2nd	C	C	C	**CD**	M
3rd	M	M	**CD**	C	**V**
4th	**V**	**CD**	**V**	M	**CD**

Cookie Dough (38) beats Vanilla (26)

Pairwise Comparison Example

Number of voters	**18**	**8**	**20**	**15**	**3**
1st	CD	**V**	**M**	**V**	C
2nd	C	C	C	CD	**M**
3rd	**M**	**M**	CD	C	**V**
4th	**V**	CD	**V**	**M**	CD

Mint Chocolate Chip (41) beats Vanilla (23)

Pairwise Comparison Example

Our tally is then,

- Chocolate (2)
- Cookie Dough (3)
- Mint Chocolate Chip (1)
- Vanilla (0)

Cookie Dough Wins!

Least Worst Defeat

- First candidates are compared using Pairwise Comparison.
- Each candidate's worst loss is recorded.
- From these losses, the candidate with the smallest margin in defeat is selected as the winner.

In essence, we are selecting the candidate who loses the best, or is the best of all bad options.

Note that a Condorcet Candidate always wins using LWD.

LWD example

Since Cookie Dough was a Condorcet Candidate (and never lost) it is the winner using LWD. Let's look at the other matchups. We will record each loser and by how much they lost.

- Chocolate v. Cookie Dough (C -2)
- Chocolate v. Mint Chocolate Chip (M -24)
- Chocolate v. Vanilla (V -18)
- Cookie Dough v. Mint Chocolate Chip (M -2)
- Cookie Dough v. Vanilla (V -12)
- Mint Chocolate Chip v. Vanilla (V -18)

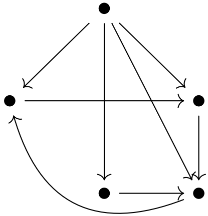
LWD example

- The worst defeat for Chocolate was -2.
- The worst defeat for Mint Chocolate Chip was -24.
- The worst defeat for Vanilla was -18.

Hence, Chocolate is second, Vanilla is third, and Mint Chocolate Chip is last.

Ranked Pairs

The method of **Ranked Pairs** uses a *directed graph* with vertices and arrows.



We will replace the vertices with the candidates.

Ranked Pairs

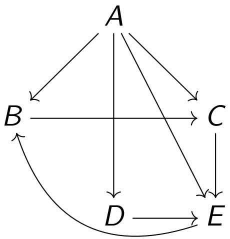
The vertex at the beginning of an arrow is called the *source* and the vertex at the end is called the *target*.

$$
\begin{tikzcd}
 & A \arrow[dl] \arrow[dr] \arrow[dd] \arrow[ddr] & \\
B \arrow[rr] & & C \arrow[d] \\
 & D \arrow[r] & E \arrow[ull, bend left]
\end{tikzcd}
$$

A is the source of four arrows but the target of no arrows. E is the target of three arrows and the source of one arrow.

Ranked Pairs

A *path* between two vertices is a collection of arrows that one may follow of successive sources and targets. A *cycle* is a path whose source and target are the same.



There is a path from A to E but no path from E to A. There is a cycle at vertex B.

Ranked Pairs

- First candidates are compared using Pairwise Comparison.
- We rank the various comparisons by margin of victory (largest to smallest).
- We begin by drawing an arrow from the winner to loser in the first victory.
- We continue *unless* a victory creates a cycle. In this case we skip the result.

The winner is the one candidate which is the source of all arrows connected to it.

Ranked Pairs Example

First we rank the outcomes of our ice cream election.

- Chocolate v. Mint Chocolate Chip (C $+24$)
- Chocolate v. Vanilla (C $+18$)
- Mint Chocolate Chip v. Vanilla (M $+18$)
- Cookie Dough v. Vanilla (CD $+12$)
- Chocolate v. Cookie Dough (CD $+ 2$)
- Cookie Dough v. Mint Chocolate Chip (CD $+ 2$)

Ranked Pairs Example

There are no cycles to worry about. Thus our ranked pairs graph is the following.

$$
\begin{tikzcd}
 & C \arrow[dr] \arrow[dd] & \\
CD \arrow[ur] \arrow[rr] \arrow[dr] & & V \\
 & M \arrow[ur] &
\end{tikzcd}
$$

Since Cookie Dough is the only vertex that is only a source (and never a target) it is the winner.

Voting Methods

We have so far discussed the following voting methods.

- Plurality Method (Majority Method)
- Borda Count Method
- Plurality-with-Elimination Method (IRV)
- Pairwise Comparison Method
- Method of Least Worst Defeat
- Ranked Pairs Method
- Approval Voting

Arrow's Impossibility Theorem

Some announcements

- Final reflections due on Monday.
- You now have all of the methods and so you can begin analyzing the results of your election.

Today's Goals

- We will discuss various fairness criteria and how the different voting methods violate these criteria.

Last Time

We have so far discussed the following voting methods.

- Plurality Method (Majority Method)
- Borda Count Method
- Plurality-with-Elimination Method (IRV)
- Pairwise Comparison Method
- Method of Least Worst Defeat
- Ranked Pairs Method
- Approval Voting

Fairness Criteria

The following fairness criteria were developed by Kenneth Arrow, an economist in the 1940s. Economists are often interested in voting theories because of their impact on *game theory*.

The mathematician John Nash (the subject of A Beautiful Mind) won the Nobel Prize in economics for his contributions to game theory.

The Majority Criterion

A majority candidate should always be the winner.

Note that this does not say that a candidate must have a majority to win, only that such a candidate should not lose.

Plurality, IRV, Pairwise Comparison, LWD, and Ranked Pairs all satisfy the Majority Criterion.

The Majority Criterion

The Borda Count Method violates the Majority Criterion.

Number of voters	6	2	3
1st	A	B	C
2nd	B	C	D
3rd	C	D	B
4th	D	A	A

Even though A had a majority of votes it only has 29 points. B is the winner with 32 points.

Moral: Borda Count punishes polarizing candidates.

The Condorcet Criterion

A Condorcet candidate should always be the winner.

Recall that a Condorcet Candidate is one that beats all other candidates in a head-to-head (pairwise) comparison.

The Condorcet Criterion

A voting method which satisfies the Condorcet Criterion is called a **Condorcet Method**.

Pairwise Comparison, LWD, and Ranked Pairs are all Condorcet Methods.

The Condorcet Criterion

The Plurality Method violates the Condorcet Criterion.

Number of voters	49	48	3
1st	R	H	F
2nd	H	S	H
3rd	F	O	S
4th	O	F	O
5th	S	R	R

R wins by the Plurality Method but loses in a head-to-head with H.

IRV and Borda Count also violate the Condorcet Criterion.

The Monotonicity Criterion

If candidate X is the winner, then X would still be the winner had a voter ranked X higher in his preference ballot.

Plurality, Borda Count, Pairwise Comparison, LWD, and Ranked Pairs all satisfy the Monotonicity Criterion.

The Monotonicity Criterion

IRV violates the Monotonicity Criterion.

Number of voters	7	8	10	2
1st	A	B	C	A
2nd	B	C	A	C
3rd	C	A	B	B

B has the fewest first place votes and is therefore eliminated. The result is that C wins.

Look at what happens if the 2 people in the last column actually rank C higher.

The Monotonicity Criterion

IRV violates the Monotonicity Criterion.

Number of voters	7	8	10	2
1st	A	B	C	**C**
2nd	B	C	A	**A**
3rd	C	A	B	B

Now A has the fewest first place votes and so it is eliminated. The result is that B wins.

Moral: IRV is vulnerable to *insincere voting*.

The Independence-of-Irrelevant-Alternatives (IIA) Criterion

If candidate X is the winner, then X would still be the winner had one or more of the irrelevant alternatives not been in the race.

Plurality, Borda Count, IRV, Pairwise Comparison, and Ranked Pairs all violate the IIA Criterion.

The IIA Criterion

Pairwise comparison violates the IIA Criterion.

Number of voters	2	6	4	1	1	4	4
1st	A	B	B	C	C	D	E
2nd	D	A	A	B	D	A	C
3rd	C	C	D	A	A	E	D
4th	B	D	E	D	B	C	B
5th	E	E	C	E	E	B	A

A has 3 wins and 1 loss. A wins with 3 points.

Now suppose C is removed (C is an irrelevant alternative).

The IIA Criterion

Pairwise comparison violates the IIA Criterion.

Number of voters	2	6	4	1	1	4	4
1st	A	B	B	B	D	D	E
2nd	D	A	A	A	A	A	D
3rd	B	D	D	D	B	E	B
4th	E	E	E	E	E	B	A

Now B has 2 wins and 1 tie but A only has 2 wins. B wins.

Arrow's Impossibility Theorem

The fairness criteria set down by Kenneth Arrow are

- The majority criterion
- The Condorcet criterion
- The monotonicity criterion
- The independence of irrelevant alternatives (IIA) criterion

Theorem

It is mathematically impossible for a voting method to satisfy all four of these fairness criteria.

Power

Some announcements

- Homework due on Wednesday.
- Projects due a week from Wednesday at the final exam.

Today's Goals

- We're going to play a game!
- This game will introduce us to weighted voting and the different potential outcomes.

Weighted Voting Game

You are a player in a voting game. There are six rounds to this game.

For each round, your vote has been assigned a weight.

There is a quota for each round. Your goal is to form a coalition of other players so that your total weight is more than the quota.

Once coalitions have formed, we will take a vote to see who gets the points for that round.

Weighted Voting Game

However, the winnings from each round will be divided amongst the coalition equally.

Therefore, it is in your best interest to form the smallest possible coalition that reaches the quota.

Terminology

- Weighted Voting System
- Motions (yes/no)
- Players $(P_1, P_2, \ldots, P_N)$
- Weights $(w_1, \ldots, w_N)$, total weight $(V = w_1 + \cdots + w_N)$
- Quota q
- Coalition

Our notation for a system will be $[q : w_1, \ldots, w_N]$.

Types of systems

Anarchy: Multiple coalitions reach the quota so we have the possibility that both yes and no can pass.

Gridlock: The quota is greater than the total weight. Hence, no motion can ever pass.

We would prefer to avoid anarchy and gridlock. Hence, we will usually only consider systems satisfying $V/2 < q \leq V$. That is, the quota lies somewhere between the total number of votes and a simple majority.

One person-one vote: When $q = V$, any motion requires the unanimous support of all players. Thus, the system is equivalent to $[N : 1, 1, \ldots, 1]$.

Types of systems

Dictator: If $w_1 \geq q$, then P_1 is a dictator and the motion passes or fails depending on P_1's decision. Every other player is irrelevant.

Veto Power: A player has veto power if its weight w satisfies

$$\begin{aligned} w < q & \ \textbf{not a dictator} \\ V - w < q & \ \textbf{remaining votes are less than the quota} \end{aligned}$$

That is, the player cannot force a motion to pass, but no motion can pass without the player's support.

Power

What is power?

Power is not just the weight that a certain player's vote has or the total weight of a coalition.

It is how much influence a player or coalition can exert compared to its total weight.

Types of coalitions

Any group of players who join forces and vote the same way is a **coalition**.

We write the coalition consisting of the first three players as $\{P_1, P_2, P_3\}$. Note that this is equivalent to any **permutation** of this set.

The coalition consisting of all players is called a **grand coalition**.

Winning/losing coalition: A coalition which has enough votes to always win/lose.

Critical players

A **critical player** in a coalition is one who is necessary in order for that coalition to win.

A player P with weight w is critical in a coalition with total weight W if $W - w < q$.

The number of winning coalitions in which a player is critical is referred to as that player's **critical count**.

A player is **irrelevant** when its critical count is 0.

Critical count

Consider the weighted voting system $[6 : 4, 3, 2, 1]$ and label the players A, B, C, D.

The possible winning coalitions are:

$$\{A, B, C, D\}, \{A, B, C\}, \{A, B, D\}, \{A, C, D\}, \{B, C, D\}, \{A, B\}, \{A, C\}.$$

The critical players in each coalition are:

$$\{A, B, C, D\}, \{\mathbf{A}, B, C\}, \{\mathbf{A}, \mathbf{B}, D\}, \{\mathbf{A}, \mathbf{C}, D\}, \{\mathbf{B}, \mathbf{C}, \mathbf{D}\}, \{\mathbf{A}, \mathbf{B}\}, \{\mathbf{A}, \mathbf{C}\}.$$

The critical counts are then:

$$A = 5, B = 3, C = 3, D = 1.$$

Banzhaf Power Index

The Banzhaf Power Index is a measure of the size of a player's power.

Let $P_1, \ldots, P_N$ be the players in a weighted voting system with critical counts $B_1, \ldots, B_N$. Set $T = B_1 + \cdots + B_N$ (the total critical count).

A player's Banzhaf Power Index (BPI) is

$$\beta_i = \frac{B_i}{T}.$$

Banzhaf Power Index example

Consider the weighted voting system $[6 : 4, 3, 2, 1]$ again. We had a total critical count of $T = 12$.

The player BPIs are:

$$\beta_A = \frac{5}{12}, \quad \beta_B = \beta_C = \frac{3}{12} = \frac{1}{4}, \quad \beta_D = \frac{1}{12}.$$

Note that the sum of the BPIs will always be 1.

What is the BPI when one player is a dictator?

What is the BPI when one player has veto power?

How we pick a president

On Tuesday, November 8, 2016, we went to the polls to elect a new president.

But, as you probably remember from civics class in high school, the process for picking a president is a bit convoluted. Instead of a majority or plurality method, the US uses a version of a weighted voting system called the Electoral College.

Essentially, each state is assigned a certain number of electoral votes. In general, the plurality winner of each state wins those electoral votes. Two states (Maine and Nebraska) assign electoral votes proportionally.

The number of electoral votes a state receives is determined by the number of representatives from that state plus 2 additional, with DC receiving 3.

Electoral College

How we pick a president

There are 538 total electoral votes.

The winner of the Electoral College is decided by the Majority Method. That is, a candidate must have 270 electoral votes to win.

Ties in the electoral college are decided by congress. The House of Representatives chooses the president, with each state's delegation casting one vote. The Senate chooses the vice president.

How we pick a president

Why do we use the Electoral College?

The original plan was for congress to elect the president. Others wanted states to pick and some wanted popular vote. The Electoral College (not called that at the time) was a compromise of sorts that maintained some power for the states.

The Electoral College is a (very large) weighted voting system. There are $2^{51} - 1$ coalitions (not all winning).

To compute BPI in the Electoral College, one needs mathematical software. Thankfully, someone already did it for us. (We are ignoring the fact that Maine and Nebraska distribute their votes proportionally.)

Voter Power

How do we compute the power of a given voter?

The Banzhaf power index of an individual voter in a state with population P is roughly proportional to $\frac{1}{\sqrt{P}}$ times the Banzhaf power index of the state.

The population of North Carolina is approximately 10 million and its BPI is 2.7%. The BPI of an individual North Carolina voter is

$$\frac{1}{\sqrt{10,000,000}} \cdot .027 \approx .00085\%$$

How does this compare to a state like California (population approximately 39 million)?

Spherical Geometry

Today's Goals

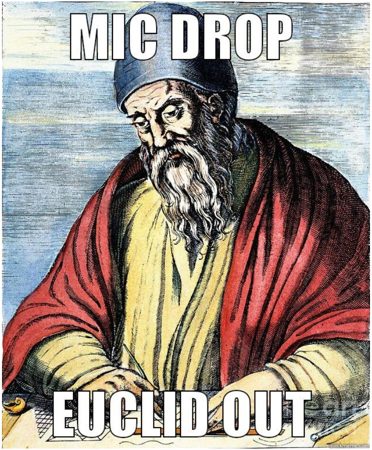
- Discuss spherical geometry and how it differs from standard (Euclidean) geometry.
- Construct triangles on spheres and analyze their properties.

Euclid

Euclid (of Alexandria) was a Greek mathematician and is known as the father of modern geometry.

Little is known about him or his life. He is believed to have lived sometime between the 4th and 3rd centuries BC.

His book, *Elements* established the foundations of modern geoemtry.



Euclid's postulates

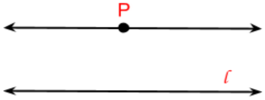
Early on in *Elements* Euclid states the following five *postulates* or axioms on which he builds the fundamentals of what we now call **Euclidean Geometry**:

1. There is a Line joining any two points.
2. Any Line segment can be extended to a Line segment of any desired length.
3. For every Line segement L, there is a Circle which has L as a radius.
4. All right angles are congruent to each other.
5. If two lines are drawn which intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines inevitably must intersect each other on that side if extended far enough.

The Parallel Postulate

The fifth postulate is equivalent to the following.

The Parallel Postulate: If P is a point and L is a line not passing through P, there is exactly one line through P parallel to L.



(This postulate is also equivalent to the Pythagorean Theorem.)

Hyperbolic Geometry

Euclid proved the first 28 propositions in *Elements* without using the parallel postulate but was forced to use it by the 29th. Many mathematicians felt that the parallel postulate should follow from the first four but none could prove this.

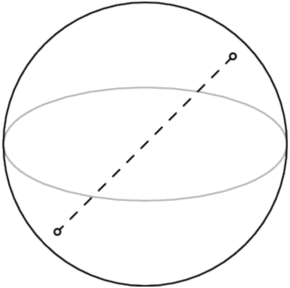
János Bolyai and Nikolai Lobachevskii independently developed a system of geometry, called Hyperbolic Geometry, that obeyed the first four axioms in Euclidean Geometry *but not* the parallel postulate. Carl Frederich Gauss had also worked out many similar ideas.

We call such geometries **Non-Euclidean Geometries**.

Spherical Geometry

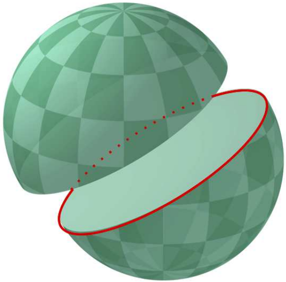
Another Non-Euclidean Geometry is known as **Spherical Geometry**.

A **Point** in Spherical Geometry is actually a pair of *antipodal points* on the sphere, that is, they are connected by a line through the center of a sphere. For example, the north and south pole of the sphere are together one point.



Spherical Geometry

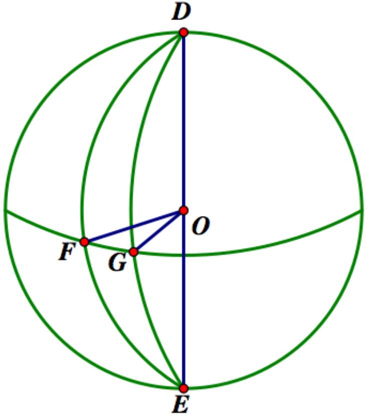
A **Line** in Spherical Geometry is a *great circle* on the sphere, that is, a circle that divides the sphere into two equal halves. For example, the equator is a great circle.



A Line (great circle) is considered infinite (actually *boundless*) because one can travel on it indefinitely.

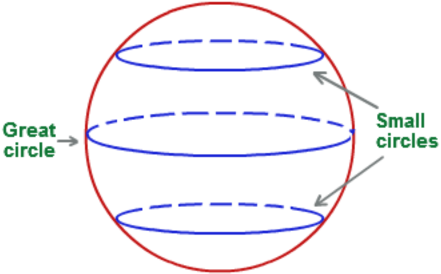
Spherical Geometry

Given any pair of (non-antipodal) Points, there is a Line (great circle) that connects them.



Spherical Geometry

A **Circle** in Spherical Geometry is just a circle and it can have any length.



Spherical Geometry

Recall that the parallel postulate states that if P is a point and L is a line not passing through P, there is exactly one line through P parallel to L.

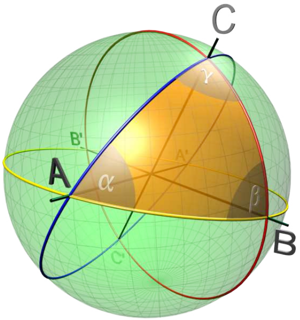
This is *false* in Spherical Geometry. It is impossible for two great circles to both divide the sphere in two equal halves and not intersect.

Moreover, any two distinct Lines (great circles) intersect in *exactly* one Point. The same thing happens in Projective Geometry, which is closely connected to Spherical Geometry.

Spherical Geometry

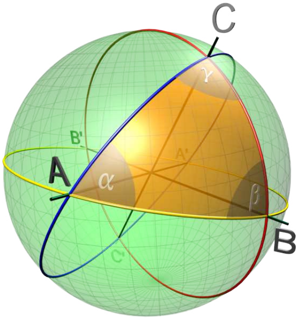
There are many other ways in which Spherical Geometry is different from Euclidean Geometry.

A **Triangle** in Spherical Geometry is formed by the intersection of three Lines (great circles) in three points (vertices).



Spherical Geometry

In Euclidean Geometry, the sum of the angles in a triangle is 180°
In Spherical Geometry, the sum of the angles in a Triangle is between 180° and 540°.



The Point

A semi-reasonable question at this point is the following:

What does this have to do with voting?

The answer is: not much, except...

All math is built upon axioms, or fundamental beliefs. This is true in voting theory, it's also true in geometry, algebra, and, for that matter, anything scientific and even things we *perceive* to be true rest on the belief that we can trust our own perceptions.

The Point

Sometimes it is worth trusting our instincts and intuition so that we can move forward. Otherwise you can get stuck, as many mathematicians did, trying to prove the axioms themselves.

On the other hand, it's also worth taking time to challenge our intuition and the axioms we build mathematics (or life) on to see what might emerge.

You might find that something new will emerge, or you might find, as Arrow did, that the axioms themselves are contradictory.

The Point

For example, you might consider yourself a staunch Conservative or Liberal and, if so, you probably surround yourself (mostly) with people who think like you. This is especially true on social media sites where people find themselves in echo chambers of their own beliefs that are constantly reinforced.

You might start to think that your beliefs are fundamental axioms and that all of society should use these as a starting place for progress.

One of the best things you can do is to challenge yourself and your beliefs by learning about "the other side". Read a book, start a conversation, or take a class about different political/social philosophies. You may not change your views, but you'll open yourself up to understanding society, and other people, a bit better.