Lecture 3

Aggregation and Least Squares

History of Data Science, Spring 2022 @ UC San Diego

Suraj Rampure

Announcements

- Homework 3 is released and is due on **Sunday, April 17th at 11:59PM**.
- Homework 1 is graded! Make sure to look at the solutions, posted on Slack and on the course website.

Agenda

- Pythagorean means.
- Tycho Brahe’s use of the mean.
- A pre-cursor to least squares – Boscovich’s method.
- Legendre, Gauss, and least squares.

Means

Means

- The concept of the "arithmetic mean" was known to the Pythagoreans – in fact, they are known for establishing three types of means.
- However, means were not used for the purposes of **summarizing data** until much, much later.

Pythagorean means

mean a, b

$a \geq c \geq b$

From Archytas (member of the Pythagorean school of thought)[1]:

"There are three 'means' in music: one is the arithmetic, the second is the geometric, and the third is the subcontrary, which they call 'harmonic'. The arithmetic mean is when there are three terms showing successively the same excess: the second exceeds the third by the same amount as the first exceeds the second. In this proportion, the ratio of the larger numbers is less, that of the smaller numbers greater."

$c - b = a - c$

$2c = a + b$

$$\boxed{c = \frac{a+b}{2}}$$

General: $\frac{a_1 + a_2 + \dots + a_n}{n}$

1. http://www.cs.uni.edu/~campbell/stat/pyth.html

Pythagorean means

$$c = \frac{2}{\frac{1}{a} + \frac{1}{b}}$$

From Archytas (member of the Pythagorean school of thought)[1]*:*

"The geometric mean is when the second is to the third as the first is to the second; in this, the greater numbers have the same ratio as the smaller numbers."

c b a c

$$\frac{c}{b} = \frac{a}{c} \Rightarrow c^2 = ab \Rightarrow c = \sqrt{ab}$$

General: $(a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n)^{\frac{1}{n}}$

"The subcontrary, which we call harmonic, is as follows: by whatever part of itself the first term exceeds the second, the middle term exceeds the third by the same part of the third. In this proportion, the ratio of the larger numbers is larger, and of the lower numbers less."

$$\frac{a-c}{a} = \frac{c-b}{b} \Rightarrow 1 - \frac{c}{a} = \frac{c}{b} - 1$$

$$2 = c\left(\frac{1}{a} + \frac{1}{b}\right)$$

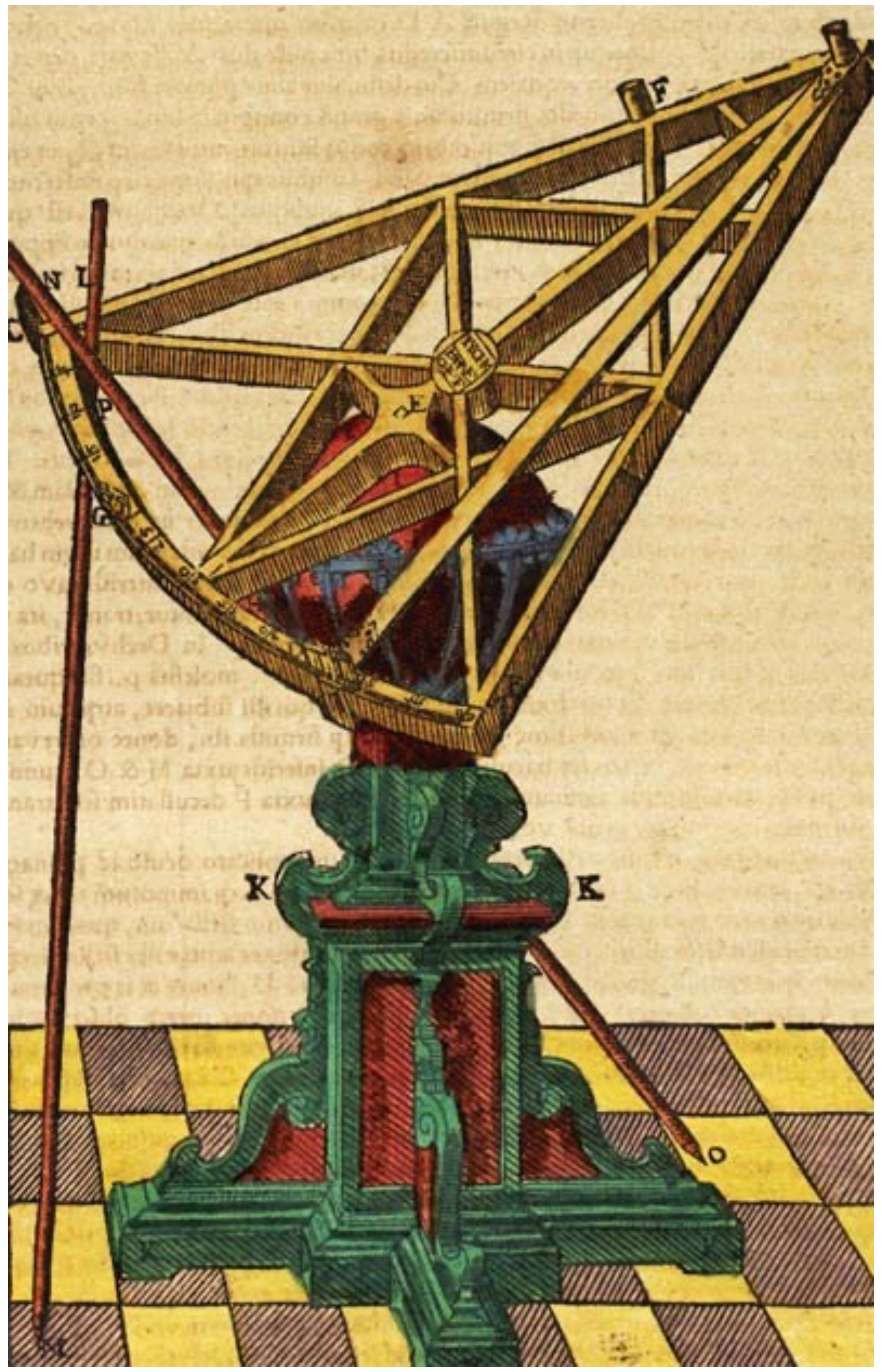
1. http://www.cs.uni.edu/~campbell/stat/pyth.html

General harmonic mean:

$$\frac{n}{\frac{1}{a_1}+\frac{1}{a_2}+\cdots+\frac{1}{a_n}}$$

Tycho Brahe

- Recall, **Tycho Brahe** (1546-1601) was a Danish astronomer.[1]
- He was a pioneer in measuring the positions of stars in the night sky, without the use of telescopes.
- Kepler used Brahe's data when creating his laws of planetary motion.
- He is also one of the earliest scientists documented as having used the mean to **combine observations**.[2]
- Also supposedly lost his nose in a fight and wore a fake nose.

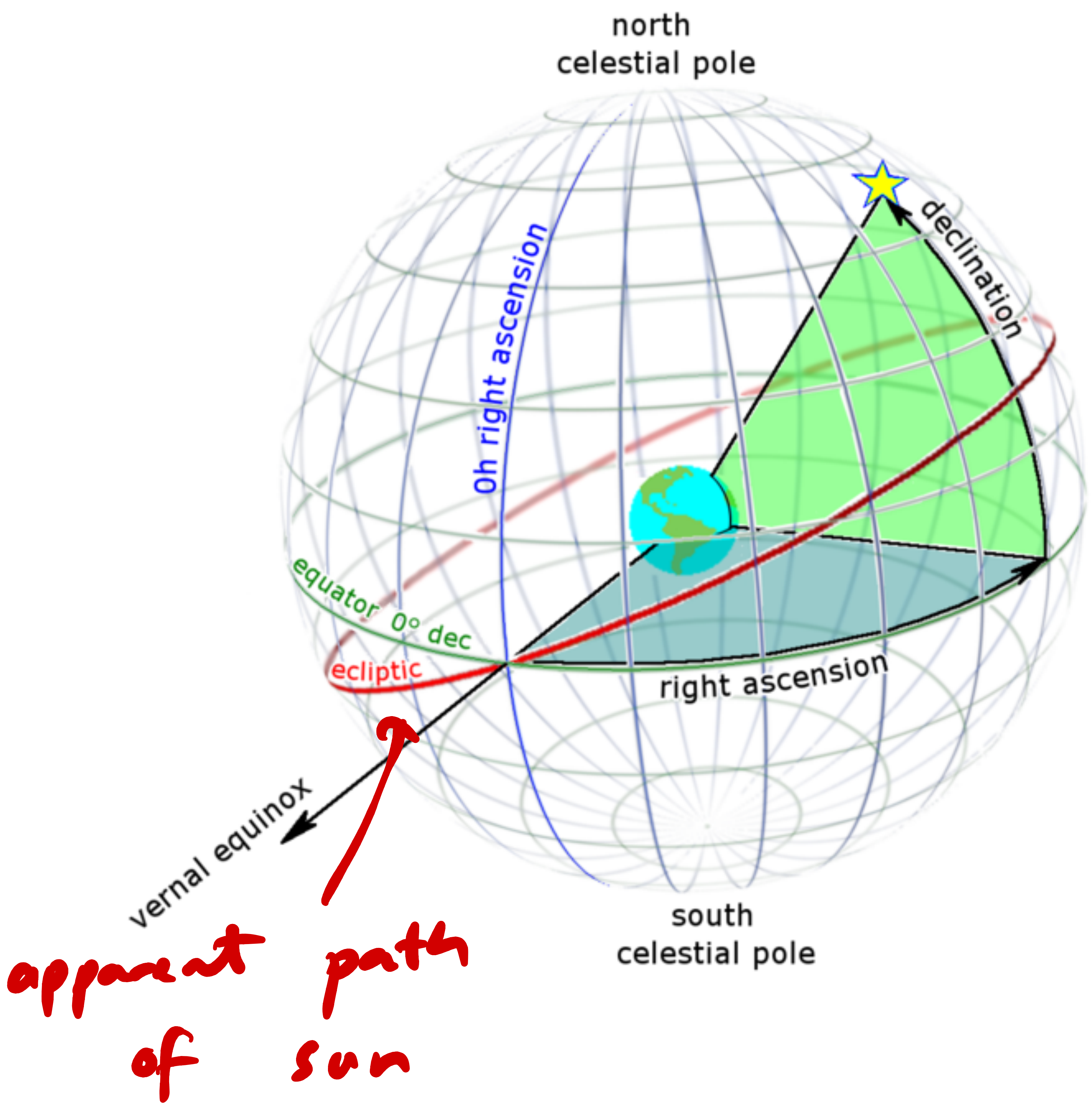


Tycho Brahe's triangular sextant

1. https://www.britannica.com/biography/Tycho-Brahe-Danish-astronomer
2. Pearson and Kendall, *Studies in the History of Probability and Statistics*, p122-123

Right ascension

- One of the earliest documented examples of **combining observations** is in the work of Tycho Brahe, who was measuring the **right ascension** of α Arietis (a star).
  - **Right ascension** is the celestial equivalent of **longitude** on Earth.
  - It is measured in units of **time**, relative to when a reference point (the "vernal equinox") passes overhead.
    - e.g. if an object's right ascension is 2 hours and 15 minutes, you will see it pass directly above you 2 hours and 15 minutes after the reference point does.
    - Similar to GMT-8 meaning "8 hours before Greenwich Meridian Time."



1582 February 26		26° 0′ 44″
1582 March 20		26 0 32
1582 April 3		26 0 30
1582 February 27	26° 4′ 16″	26 0 20
1585 September 21	25 56 23	
1582 March 5	25 56 33	26 0 38
1585 September 14	26 4 43	
1582 March 5	25 59 15	26 0 18
1585 September 15	26 1 21	
1582 March 9	25 59 49	26 0 32
1585 September 15	26 1 16	
1586 December 26	25 54 51	26 0 42
1588 December 15	26 6 32	
1586 December 27	25 52 22	26 0 37
1588 November 29	26 8 52	
1587 January 9	26 2 5	26 0 27
1588 December 6	25 58 49	
1587 January 24	26 6 44	26 0 29
1588 October 26	25 54 13	
1587 August 17	26 5 40	26 0 14
1588 April 16	25 54 48	
1587 August 17	26 1 1	26 0 4
1588 April 16	25 59 6	
1587 August 18	25 54 35	26 0 28
1588 March 28	26 6 20	
1587 August 18	25 54 49	26 0 39
1588 April 16	26 6 30	

- Brahe collected several measurements for the right ascension of α Arietis from 1582-1588, with the goal of coming up with a single value.
- He selected 3 values from 1582, and 12 values from the next 6 years, each of which was the **mean** of two other observations.
- **Question:** how do we interpret these numbers and verify that he did indeed take the mean of each pair?

Source: Pearson and Kendall, *Studies in the History of Probability and Statistics*, p122-123

Aside: measuring time in degrees

$\frac{360^\circ}{24} = \frac{24}{24}$ hours

$\frac{15^\circ}{15} = \frac{1}{15}$ hour

- Right ascension is measured in time, and can vary from 0 hours to 24 hours (because one rotation of the Earth takes 24 hours).

$\boxed{1^\circ = 4 \text{ minutes}}$

- A circle has 360° degrees in it, so one way of describing time is as using

$$\boxed{\mathbf{360^\circ = 24\ hours}}$$

- This means that **15° = 1 hour**, and **1° = 4 minutes**.

- We can further subdivide each degree into 60 **arcminutes**, denoted by ′, and each arc minute into 60 **arcseconds**, denoted by ″.

- As an example, let's try and convert the following measurement into regular minutes:

$1' = \left(\frac{1}{60}\right)^\circ$, $1'' = \left(\frac{1}{60}\right)'$ $\xrightarrow{82^\circ\ 15'\ 10''}$ $= \left(\frac{1}{3600}\right)^\circ$

$1^\circ = 4$ minutes

$1' = \left(\frac{1}{60}\right)^\circ$

$1'' = \left(\frac{1}{3600}\right)^\circ$

82° 15′ 10″

① Convert to degrees

$$82^\circ\ 15'\ 10''$$

$$= \left(82 + \frac{1}{60}\cdot 15 + \frac{1}{60^2}\cdot 10\right)^\circ$$

$$= \left(82 + \frac{1}{4} + \frac{1}{360}\right)^\circ$$

② Convert to minutes

$$\Rightarrow 4\cdot\left(82 + \frac{1}{4} + \frac{1}{360}\right) \text{ minutes}$$

$$= 328 + 1 + \frac{1}{90} = 329 + \frac{1}{90} \text{ minutes}$$

$$\boxed{5h\ 29m + \frac{1}{90}\text{ min}}$$

$$34^\circ \quad 13' \quad 4''$$

convert to regular minutes

1. Convert to degrees

$$\Rightarrow 34 + \frac{13}{60} + \frac{4}{60^2}$$

2. Convert to minutes

$$\Rightarrow 4\left(34 + \frac{13}{60} + \frac{4}{60^2}\right)$$

$$= 136 + \frac{13}{15} + \left(\frac{1}{15}\right)^2$$

$$= 2 \text{ hours, } 16 \text{ minutes} + \left[\frac{13}{15} + \left(\frac{1}{15}\right)^2\right] \text{min}$$

Back to Brahe's data

1582 February 26		26° 0′ 44″
1582 March 20		26 0 32
1582 April 3		26 0 30
1582 February 27	26° 4′ 16″	26 0 20
1585 September 21	25 56 23	
1582 March 5	25 56 33	26 0 38
1585 September 14	26 4 43	
1582 March 5	25 59 15	26 0 18
1585 September 15	26 1 21	
1582 March 9	25 59 49	26 0 32
1585 September 15	26 1 16	
1586 December 26	25 54 51	26 0 42
1588 December 15	26 6 32	
1586 December 27	25 52 22	26 0 37
1588 November 29	26 8 52	
1587 January 9	26 2 5	26 0 27
1588 December 6	25 58 49	
1587 January 24	26 6 44	26 0 29
1588 October 26	25 54 13	
1587 August 17	26 5 40	26 0 14
1588 April 16	25 54 48	
1587 August 17	26 1 1	26 0 4
1588 April 16	25 59 6	
1587 August 18	25 54 35	26 0 28
1588 March 28	26 6 20	
1587 August 18	25 54 49	26 0 39
1588 April 16	26 6 30	

- Now that we know how to interpret these numbers, we can verify that the operation Brahe used on each pair was the mean.
- Strategy: to compute mean(d_1, d_2):
  - Convert d_1 and d_2 to minutes (i.e. regular numbers) and compute their mean.
  - Convert the mean back into degrees-arcminutes-arcseconds.
- Let's try this in a **Jupyter Notebook**!

Reducing observational error

1582 February 26		26° 0′ 44″
1582 March 20		26 0 32
1582 April 3		26 0 30
1582 February 27	26° 4′ 16″	26 0 20
1585 September 21	25 56 23	
1582 March 5	25 56 33	26 0 38
1585 September 14	26 4 43	
1582 March 5	25 59 15	26 0 18
1585 September 15	26 1 21	
1582 March 9	25 59 49	26 0 32
1585 September 15	26 1 16	
1586 December 26	25 54 51	26 0 42
1588 December 15	26 6 32	
1586 December 27	25 52 22	26 0 37
1588 November 29	26 8 52	
1587 January 9	26 2 5	26 0 27
1588 December 6	25 58 49	
1587 January 24	26 6 44	26 0 29
1588 October 26	25 54 13	
1587 August 17	26 5 40	26 0 14
1588 April 16	25 54 48	
1587 August 17	26 1 1	26 0 4
1588 April 16	25 59 6	
1587 August 18	25 54 35	26 0 28
1588 March 28	26 6 20	
1587 August 18	25 54 49	26 0 39
1588 April 16	26 6 30	

- The values in the right-most column are far less spread out than the values in the middle column.
- As such, Brahe used the **mean to eliminate systematic errors**.[1]
- The final right ascension that Brahe reported was 26° 0′ 30″, which is very close to both the mean of all 15 numbers in the right column and the mean of just the bottom 12.
- Per his biographer[1], the correct value of the right ascension of α Aries at the time was 26° 0′ 45″, which is quite close.

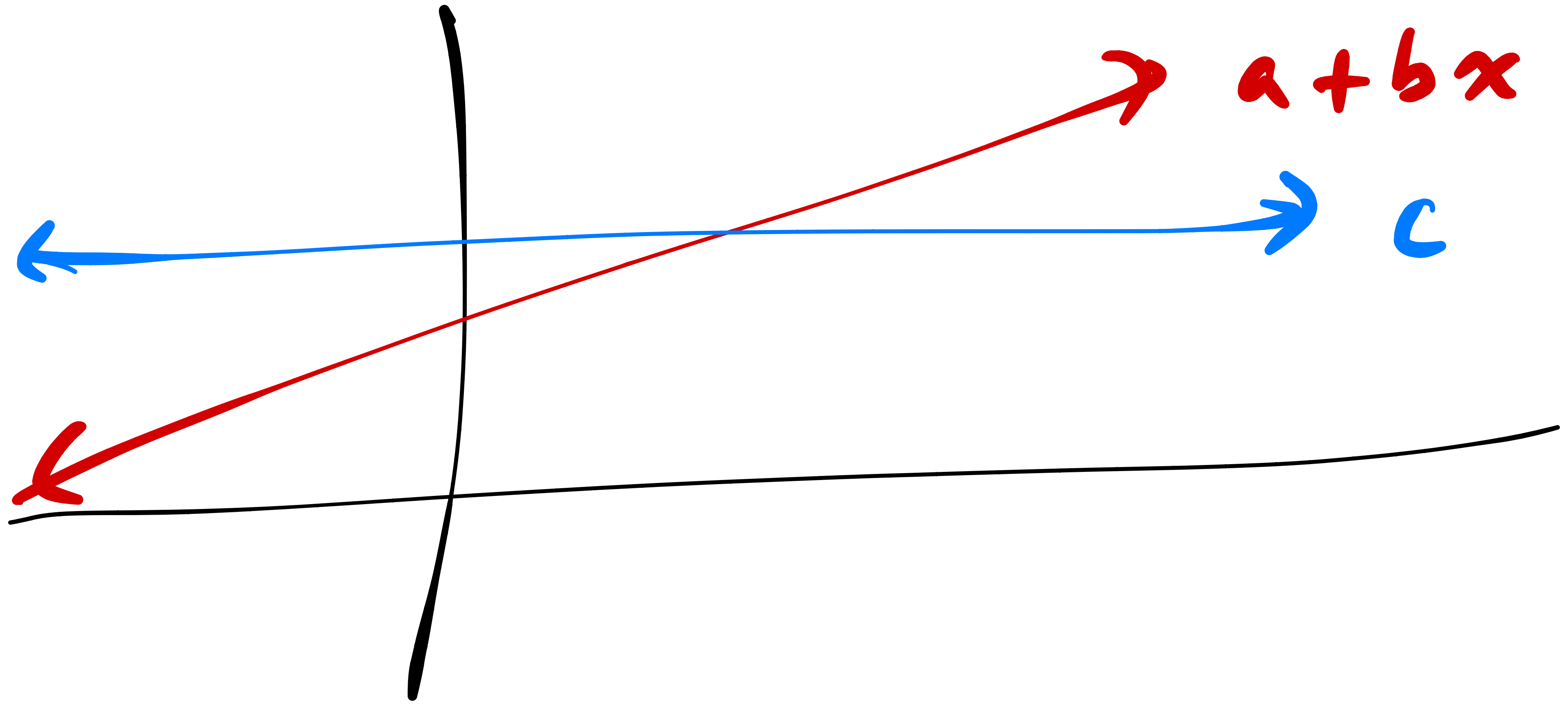
1. Pearson and Kendall, *Studies in the History of Probability and Statistics*, p122-123

The mean and least squares

For context…

- Without proper context, it may not be clear what **aggregation** (e.g. taking the mean or median of a set of values) has anything to do with **least squares** (which you learned in DSC 10 is the foundation of linear regression).
    - This connection is made more clear in DSC 40A.
- We'll spend a little bit of time providing this context, as we move into the origins of least squares.

Making predictions



- As you've seen in DSC 10, the slope and intercept of the **line of best fit** come from finding the values of a and b that minimize **mean squared error.**

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \underbrace{(a + bx_i)}\right)^2$$

- What if we want to use a more simple prediction technique – what if we want to make a **constant** prediction, for each observation?
  - To do this, we'd need to find the constant c that minimizes mean squared error.

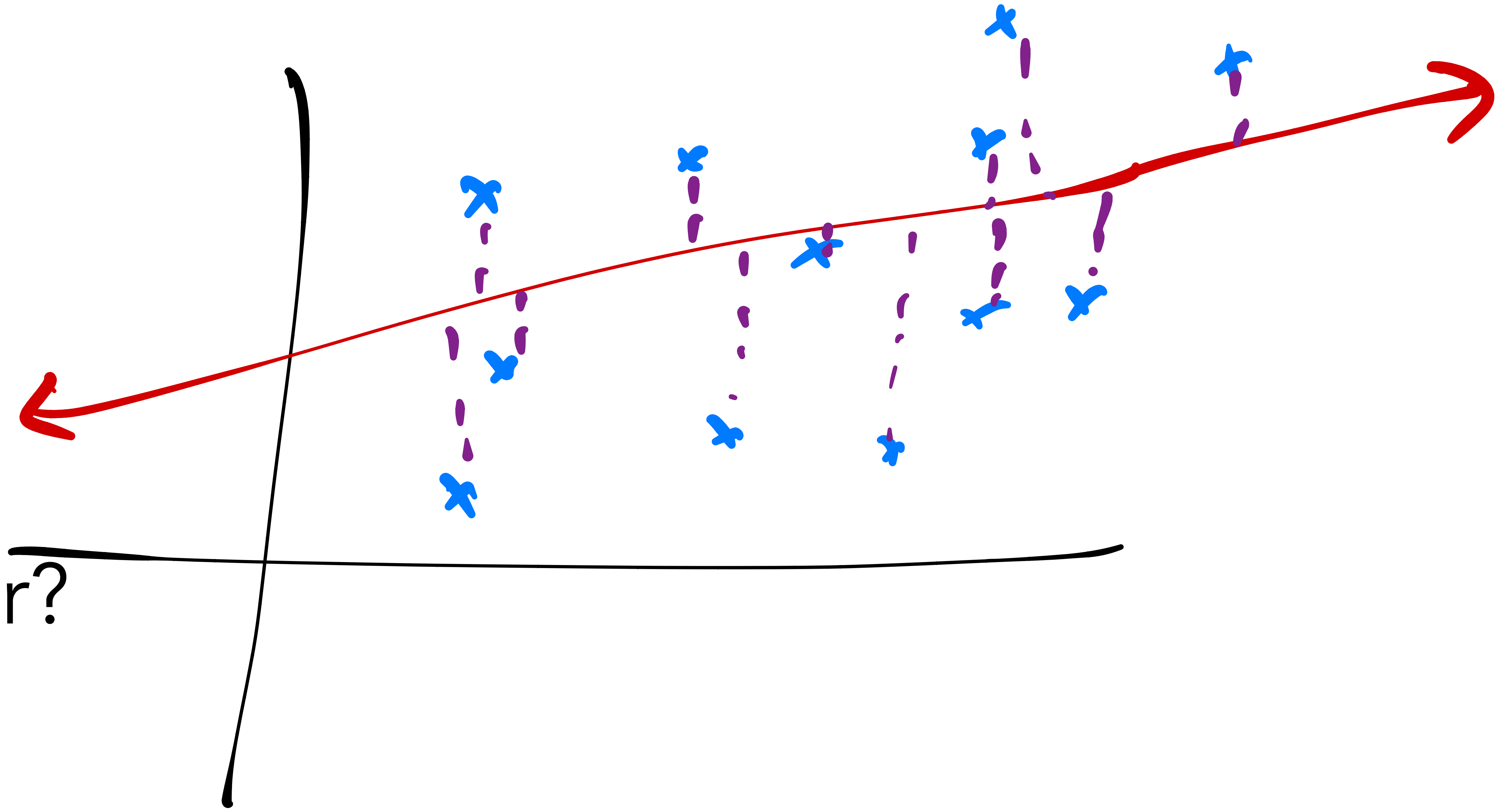
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \underbrace{c})^2$$

$$R(c) = \frac{1}{n}\sum_{i=1}^{n}(y_i - c)^2$$ → mean squared error

"least squares"

$$\boxed{c^* = \frac{y_1 + y_2 + \ldots + y_n}{n}}$$ → aggregation

Other types of error



- Why do we minimize mean **squared** error?

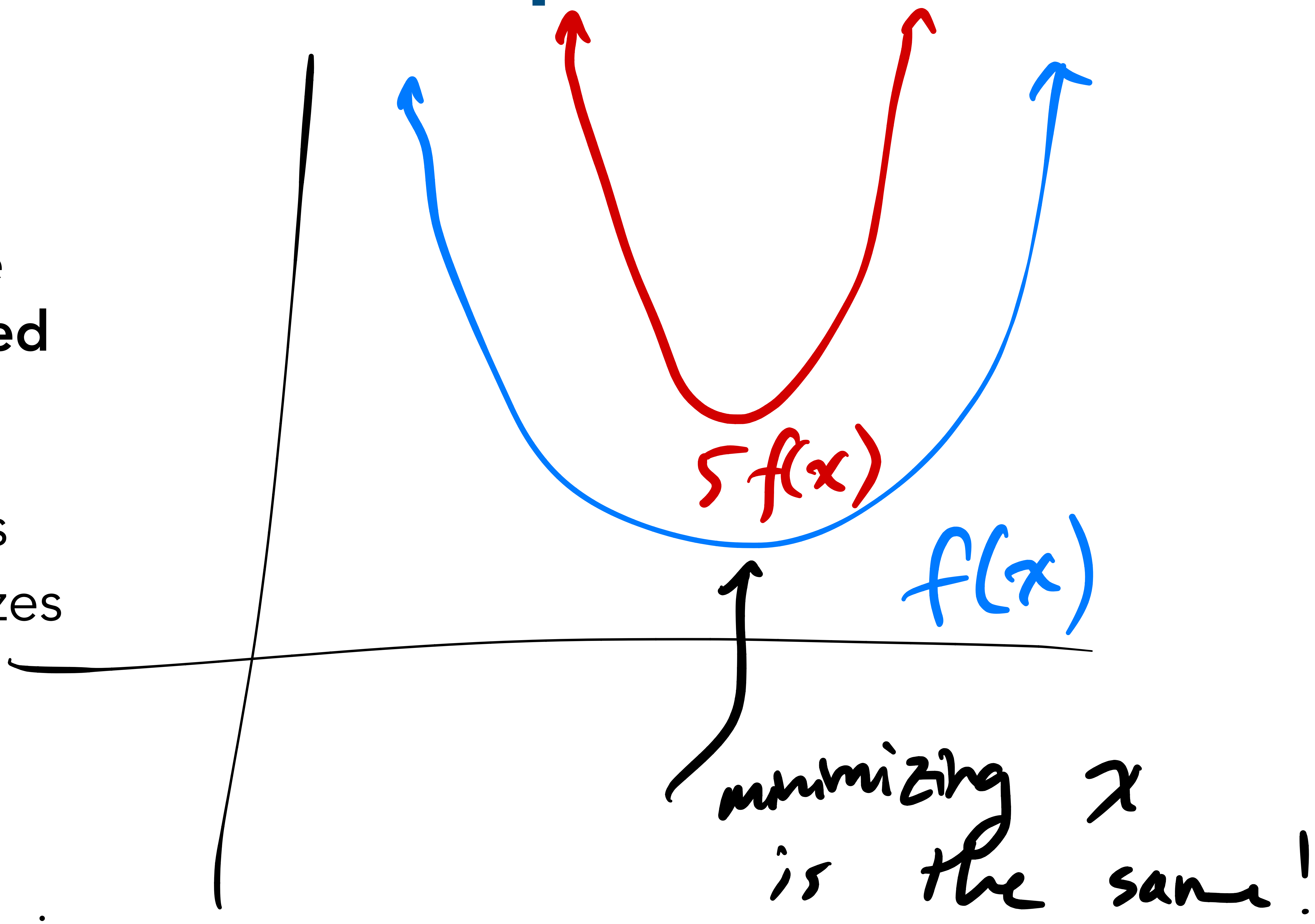
- Instead of squaring the errors before taking the mean, is there **another** operation we could apply?

could've used absolute value

mean absolute error : $\frac{1}{n} \sum_{i=1}^{n} |y_i - c|$

Mean squared error vs. sum of squared errors

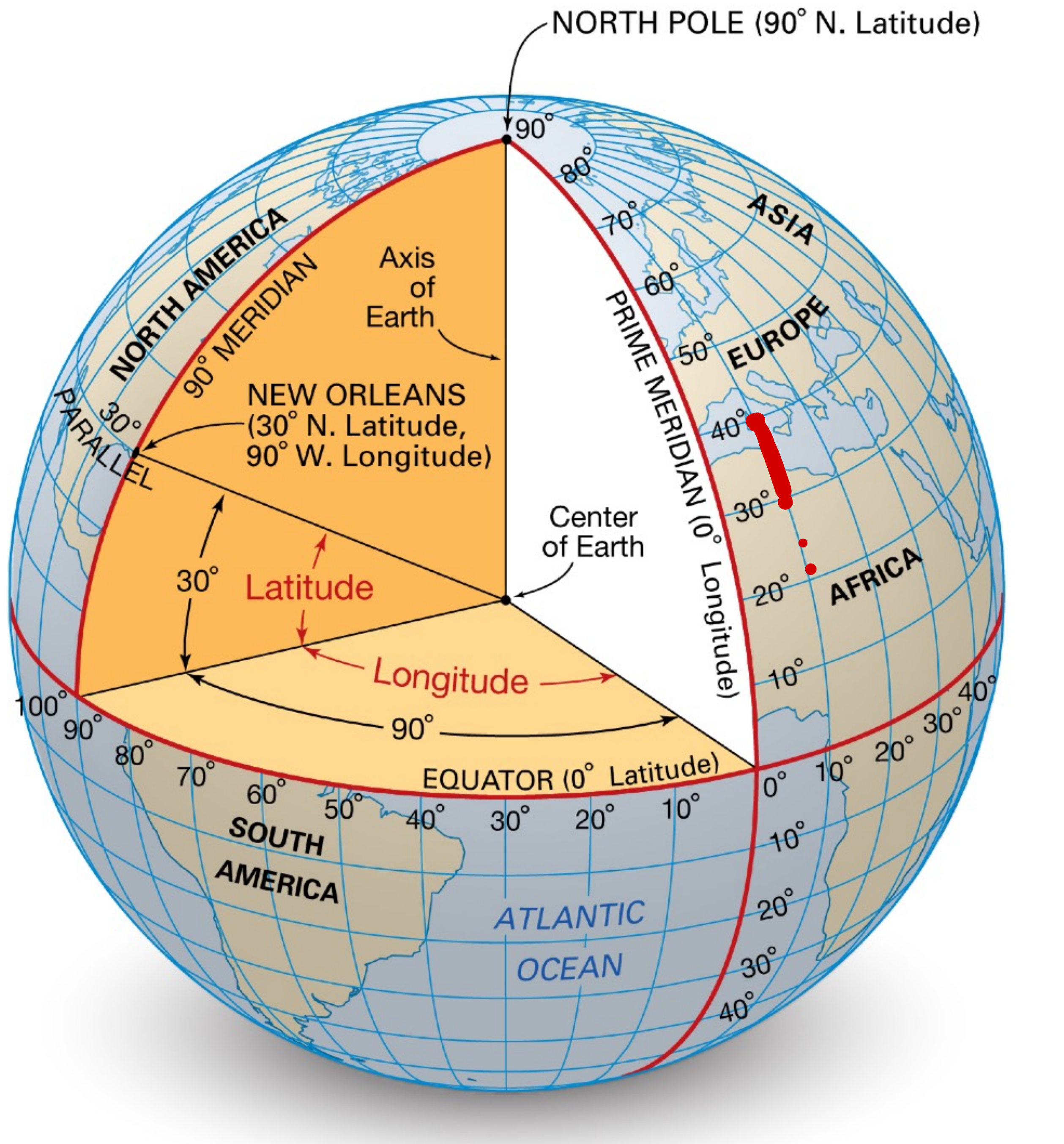
- Minimizing **mean squared error** is the same as minimizing the **sum of squared errors**.
- Key idea: the value of x that minimizes $f(x)$ is the same value of x that minimizes $c \cdot f(x)$, if c is some positive constant.
- Many of the original authors we will study aimed to minimize the sum of squared errors, not the mean – but this is the same task.



sum = $n \cdot$ mean

Boscovich's method

The length of a meridian arc



- A **meridian arc** is a curve drawn between two points on the surface of the Earth that have the same longitude.
- In the mid-1700s, **geodesists** were concerned with studying the shape of Earth.
  - Earth is an ellipsoid that is slightly flatter at the poles than it is at the equator.
- Their goal at the time was to determine the relationship between **the length of one degree of latitude near the North Pole** and **the length of one degree of latitude elsewhere on Earth**.
- To do this, they measured the lengths of several meridian arcs.

Boscovich's data

Croatia

- Roger Joseph Boscovich (1711-1787) was a Dalmatian astronomer, mathematician, and Jesuit priest.

- He obtained data containing the length of one degree of latitude at five different spots on Earth.

Table 1.4. Boscovich's data on meridian arcs.

Ecuador

Location	Latitude (θ)	Arc length (toises)	Boscovich's $\sin^2\theta \times 10^4$
(1) Quito	0°0′	56,751	0
(2) Cape of Good Hope	33°18′	57,037	2,987
(3) Rome	42°59′	56,979	4,648
(4) Paris	49°23′	57,074	5,762
(5) Lapland	66°19′	57,422	8,386

Finland

Source: Boscovich and Maire (1755, p. 500). Reprinted in Boscovich and Maire (1770, p. 482).

Note: Arc lengths are given as toises per degree measured, where 1 toise ≅ 6.39 feet. The value for $\sin^2\theta \times 10^4$ for the Cape of Good Hope is erroneous and is evidently based on 33°8′. The correct figure would be 3,014.

Source: Stigler, *Studies in the History of Probability and Statistics*, p. 43

The model

a : arc length (known)
θ : latitude (known)
z, y : unknowns

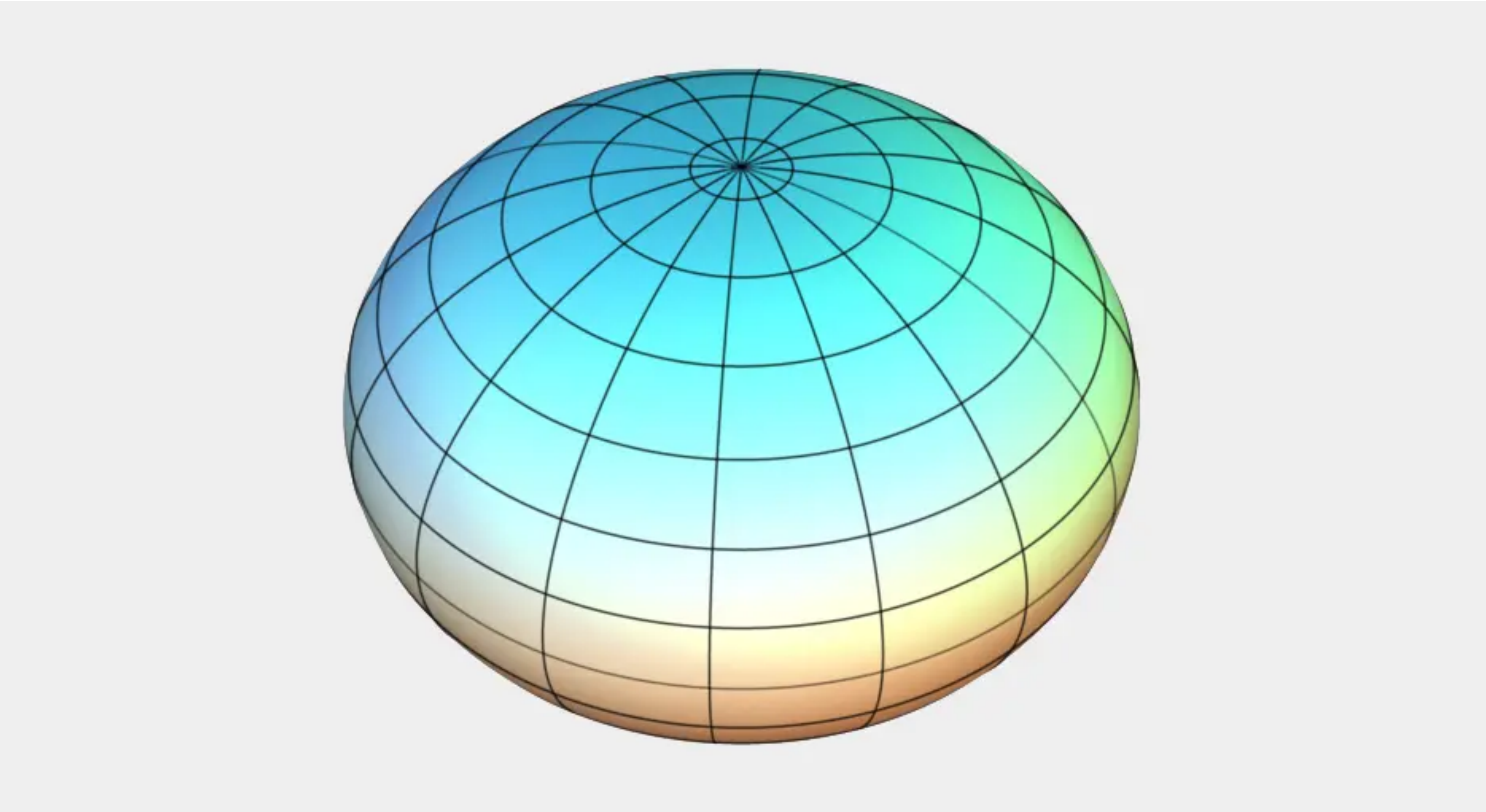
- A rough approximation for the length of an arc is

$$a = z + y \sin^2 \theta$$

  where z is the length of a degree at the equator and y is the "excess".

- If $y = 0$, then the Earth is a perfect sphere, and meridian arcs are of the same length (z) at any latitude.

- If $y > 0$, the Earth is flatter towards the poles, and meridian arcs range from length z at the equator to length $z + y$ at the North Pole.



Source

An abundance of data

$$a = z + y \sin^2 \theta$$

- If Boscovich had just 2 observations, he'd have a system of two equations and two unknowns, and would be able to solve for z and y.
- However, he had 5 observations, and had to deduce a method of computing z and y using all 5 observations.
- Ideas?

Table 1.4. Boscovich's data on meridian arcs.

Location	Latitude (θ)	Arc length (toises)	Boscovich's $\sin^2 \theta \times 10^4$
(1) Quito	0°0′	56,751	0
(2) Cape of Good Hope	33°18′	57,037	2,987
(3) Rome	42°59′	56,979	4,648
(4) Paris	49°23′	57,074	5,762
(5) Lapland	66°19′	57,422	8,386

Source: Boscovich and Maire (1755, p. 500). Reprinted in Boscovich and Maire (1770, p. 482).

Note: Arc lengths are given as toises per degree measured, where 1 toise ≅ 6.39 feet. The value for $\sin^2 \theta \times 10^4$ for the Cape of Good Hope is erroneous and is evidently based on 33°8′. The correct figure would be 3,014.

Boscovich's method

- For each of our five observations (θ_i, a_i), we can write

$$a_i = z + y \sin^2 \theta_i$$

- Boscovich's described a method for selecting z and y:

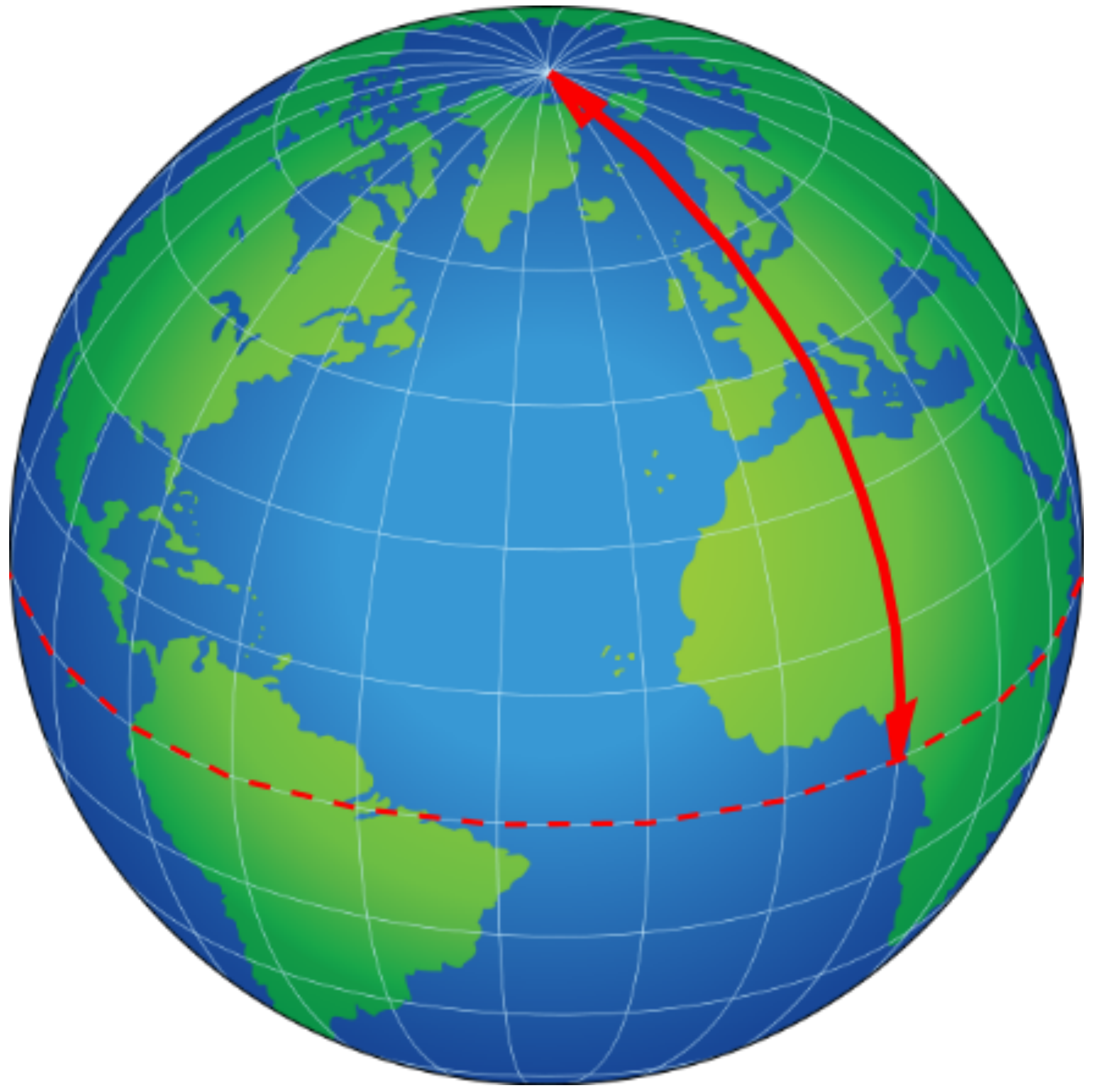
sum of absolute errors

1. For each i, write $e_i = a_i - z - y \sin^2 \theta_i$.
2. Choose z and y such that $\underbrace{\sum_i e_i = 0}_{\text{constraint}}$ and $\overbrace{\sum_i |e_i|}^{\text{sum of absolute errors}}$ is minimized.

- **What does this resemble?**

Least squares

Legendre



- Adrien-Marie Legendre (1752-1833) was a French mathematician who was also active in the field of geodesy[1].
    - In 1791, the French Academy of Science defined a meter as being one **ten millionth** of the length of the meridian arc starting at the North Pole, passing through Paris, and ending at the equator.
    - He helped measure the length of a meter.

1. https://www.britannica.com/biography/Adrien-Marie-Legendre

Legendre's least squares

- In a 1805 paper about measuring the orbits of comets, Legendre published an appendix titled "Sur la Methode des moindres quarres", which detailed a general procedure for estimating coefficients of linear equations.
- He wrote (translated):

> *"Of all the principles which can be proposed for [making estimates from a sample], I think there is none more general, more exact, and more easy of application, than that of which we have made use… which consists of rendering the sum of the squares of the errors a minimum."*

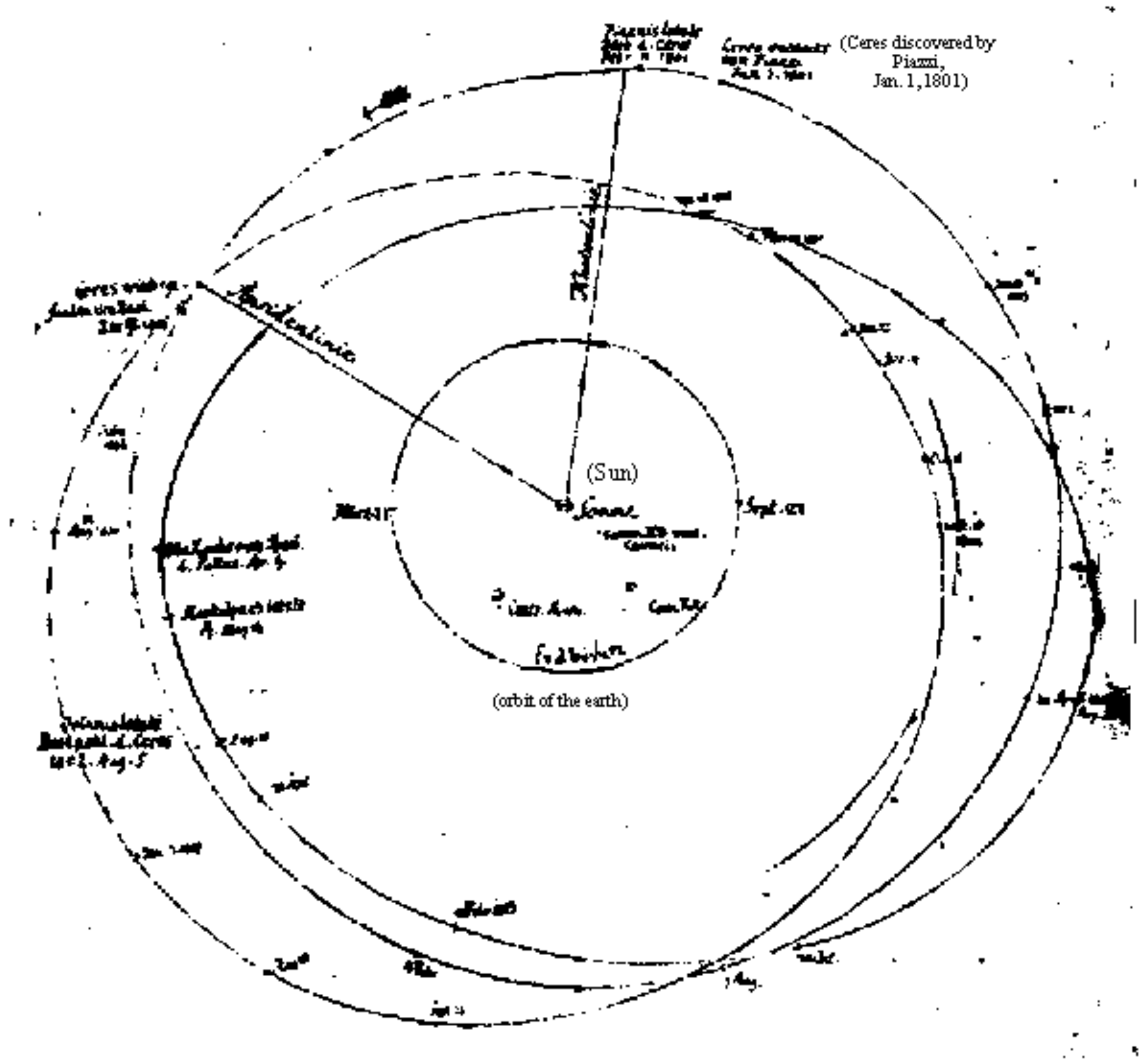
Gauss

- Carl Friedrich Gauss (1777-1855)[1] was a German mathematician, and is one of the most accomplished mathematicians of all time.
- He is known for developing or contributing to:
  - Least squares.
  - The normal (Gaussian) distribution.
  - Algebra and number theory.
    - He supposedly summed the positive integers between 1 and 100 very quickly.
  - Electromagnetism.
  - **Not** Gaussian elimination!

1. https://www.britannica.com/biography/Carl-Friedrich-Gauss

Gauss and least squares

- In **1809**, Gauss published "*Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections*", and in it he used the method of least squares to calculate the shapes of orbits.
  - Legendre published about least squares in **1805**, 4 years before. However, Gauss claimed to have known about least squares in **1795**.
  - **Evidence:** Gauss was able to predict the precise location of planetoid Ceres using his method of least squares.
    - Ceres was observed on January 1st, **1801** for a period of 40 days. Several astronomers competed to predict where it would be spotted again, and Gauss' guess was the only correct one[2].



Sketch of the orbits of Ceres and Pallas (nachlaß Gauß, Handb. 4). Courtesy of Universitätsbibliothek Göttingen.

Source

1. https://www.britannica.com/biography/Carl-Friedrich-Gauss
2. https://blog.bookstellyouwhy.com/carl-friedrich-gauss-and-the-method-of-least-squares

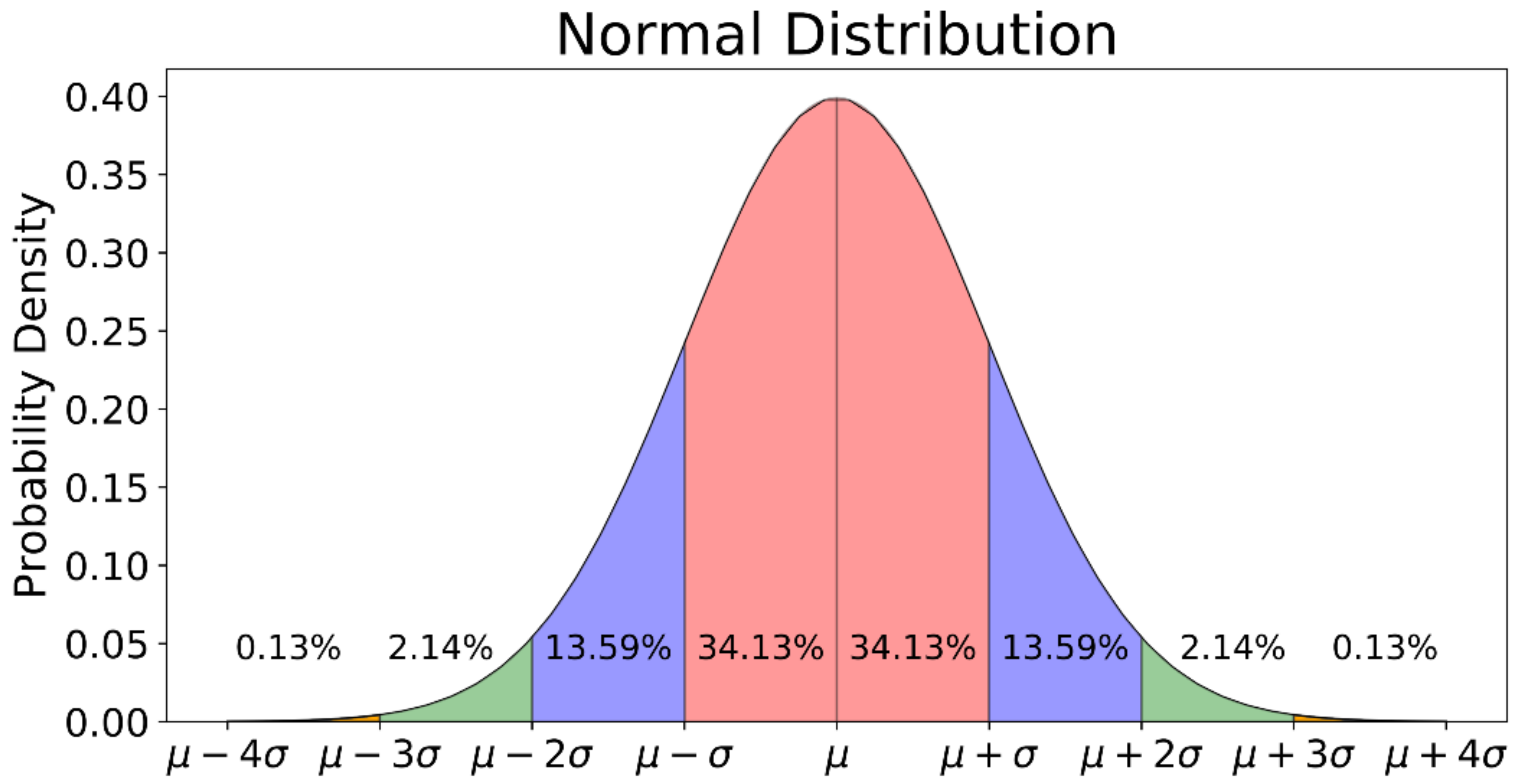
Error distributions

- One of the key differences between the approaches to least squares by Gauss and Legendre was that Gauss linked the theory of least squares to probability theory.
- Specifically, he posed the least squares **model** where

$$y_i = a + bx_i + \epsilon_i$$

where ϵ_i is a **random variable** that follows the following **error distribution**:

$$\phi(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



We will study Gauss' derivation of the (now-called) Gaussian/normal distribution in Lecture 5.

Summary, next time

Summary, next time

- Much of the advances regarding aggregation and statistical estimation in the 1500-1800s was motivated by geodesy and astronomy.
    - Tycho Brahe's use of the mean.
    - Boscovich's method regarding meridian arcs.
    - Legendre's method of least squares.
- **Next time:** Percentiles and regression.