Lecture 4

Least Squares, Regression, and Correlation

History of Data Science, Winter 2022 @ UC San Diego

Suraj Rampure

Announcements

- Classes will now be in-person! Come to class in-person (Center Hall 218) OR via Zoom (link on the course website).
    - Office hours right after lecture will be in the lecture room (Center Hall 218) + Zoom as well.
    - Friday office hours (3:30-4:30PM) will be **remote only**.
- Homework 4 will be released by tomorrow, and will be due **Sunday, February 6th at 11:59PM**.

Agenda

- Legendre and Gauss’ development of least squares.
- Quetelet and the “average man”.
- Galton’s development of regression.
- Pearson and ~~Fisher~~.

Least squares

Last time: 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. → error distribution
  - Algebra and number theory.
    - He supposedly summed the positive integers between 1 and 100 very quickly.
  - Electromagnetism. → electric potential of a field
  - **Not** Gaussian elimination!

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

$$1+2+3+ \dots + 98+99+100$$

idea: add small numbers to big numbers

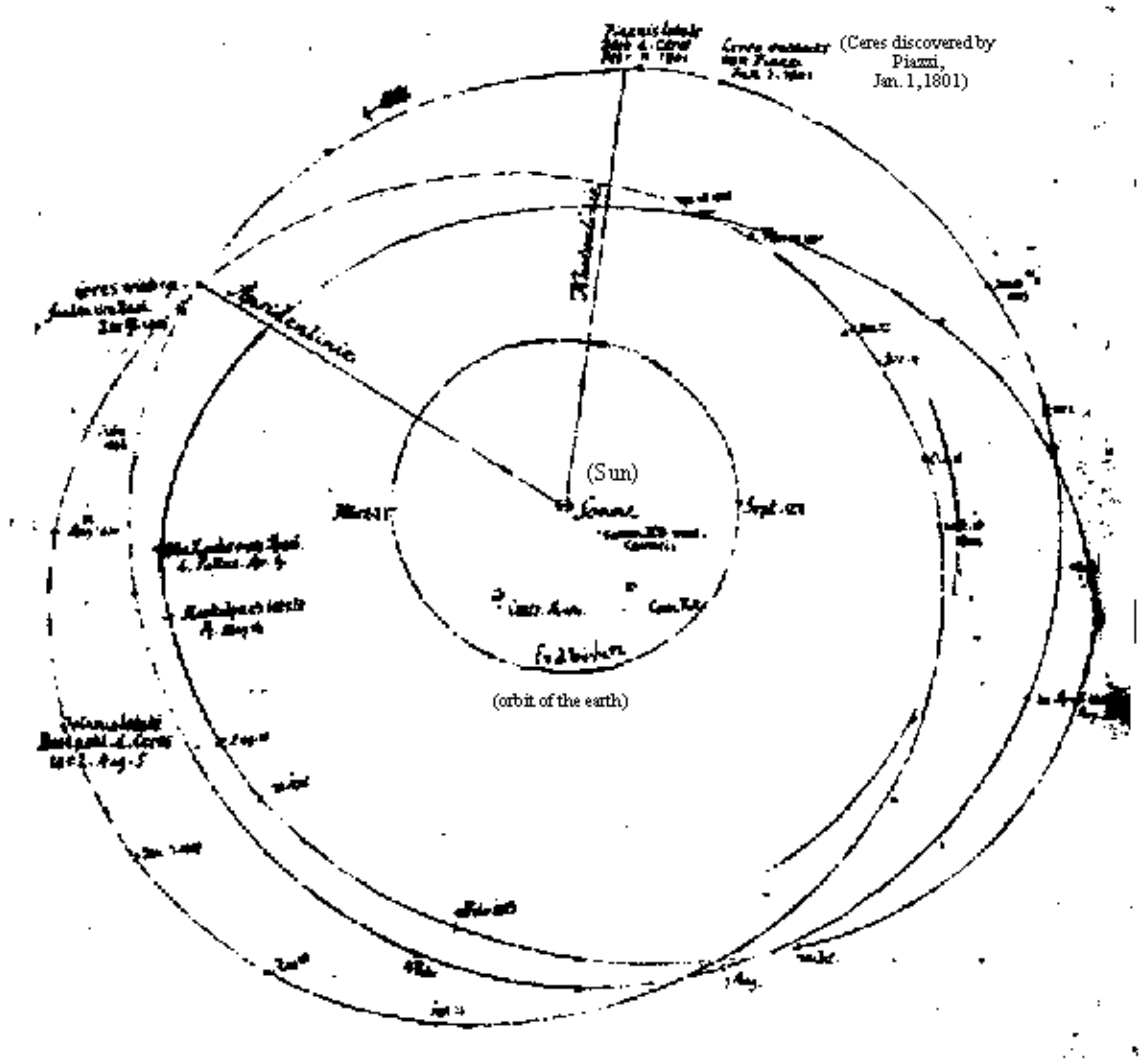
$$
\begin{array}{r}
1+2+3+4+5+\dots \\
+\;100+99+98+97+96+\dots \\
\hline
101+101+101+\dots
\end{array}
$$

$\underbrace{101+101+\dots}_{50 \text{ times}}$

$$= 101 \cdot 50 = \boxed{5050}$$

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].

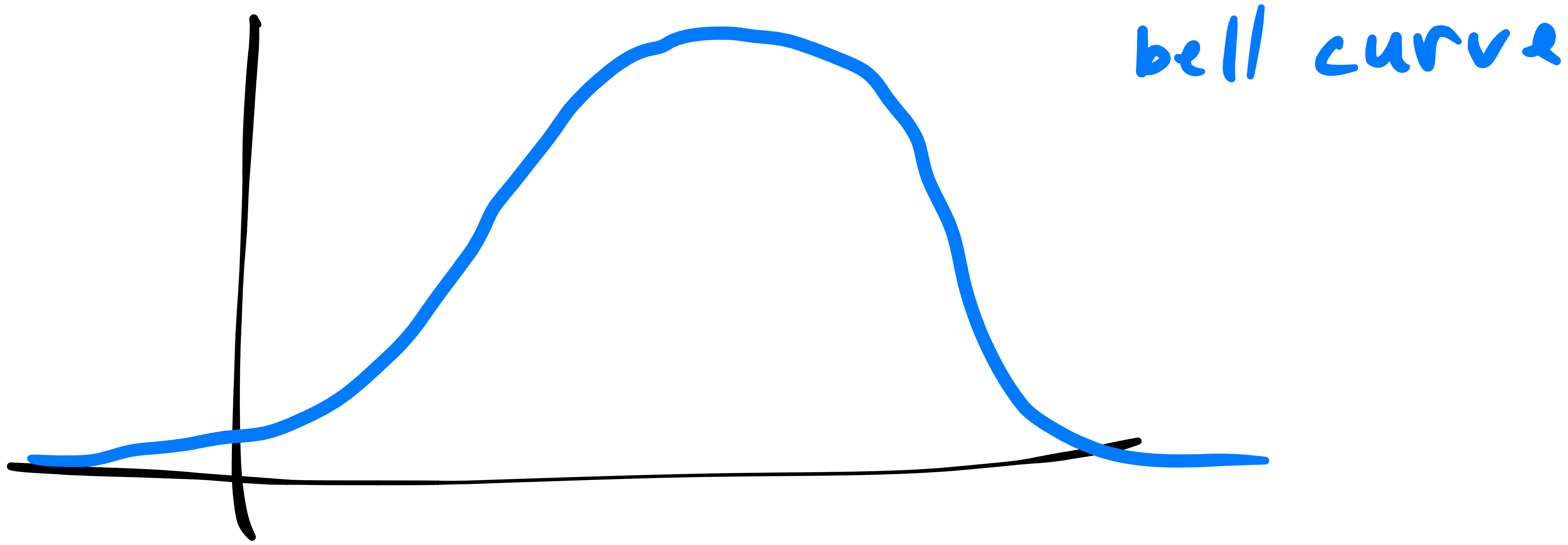


Sketch of the orbits of Ceres and Pallas (nachlaß Gauß, Handb. 4). Courtesy of Universitätsbibliothek Göttingen.

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



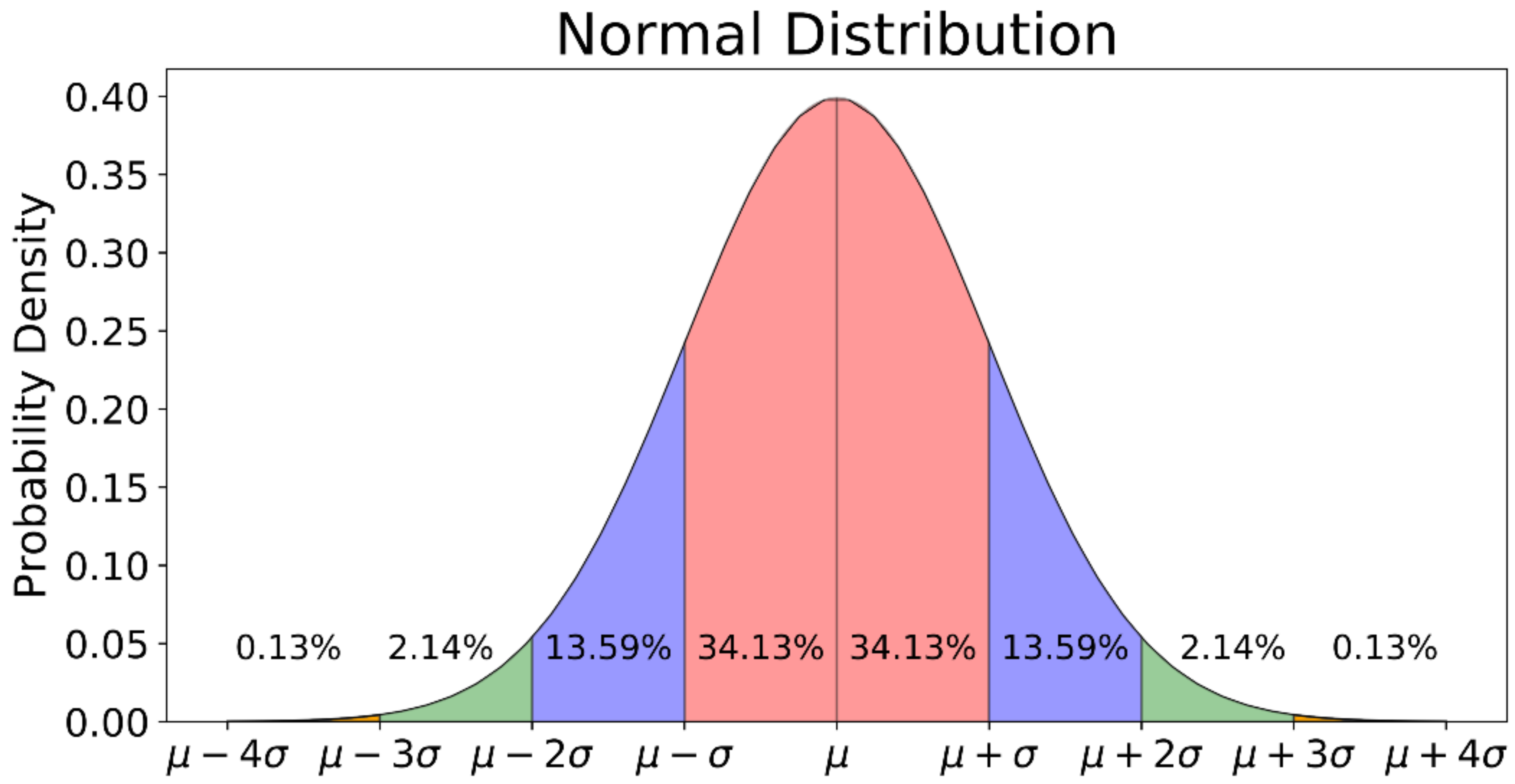
- One of the key differences between the approaches to least squares by Gauss and Legendre was that Gauss linked the theory of least squares to probability theory.
- Specifically, he posed the least squares **model** where

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

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

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

what we call Normal distribution!



We will study Gauss' derivation of the (now-called) Gaussian/normal distribution next class.

Maximum likelihood estimation (MLE)

- To illustrate the power of Gauss' contribution, we need to describe the concept of **maximum likelihood estimation**.

→ = probability it flips heads

- Suppose we have a coin, whose **bias**, p, is unknown. We flip the coin 10 times, and see the sequence HTTHTTTTTH.

- **Question: what is the best guess for the value of p?**

reasonable guess: $0.3 = \frac{3}{10}$, the fraction of heads in 10 flips

Sequence HTTHTTTTTH

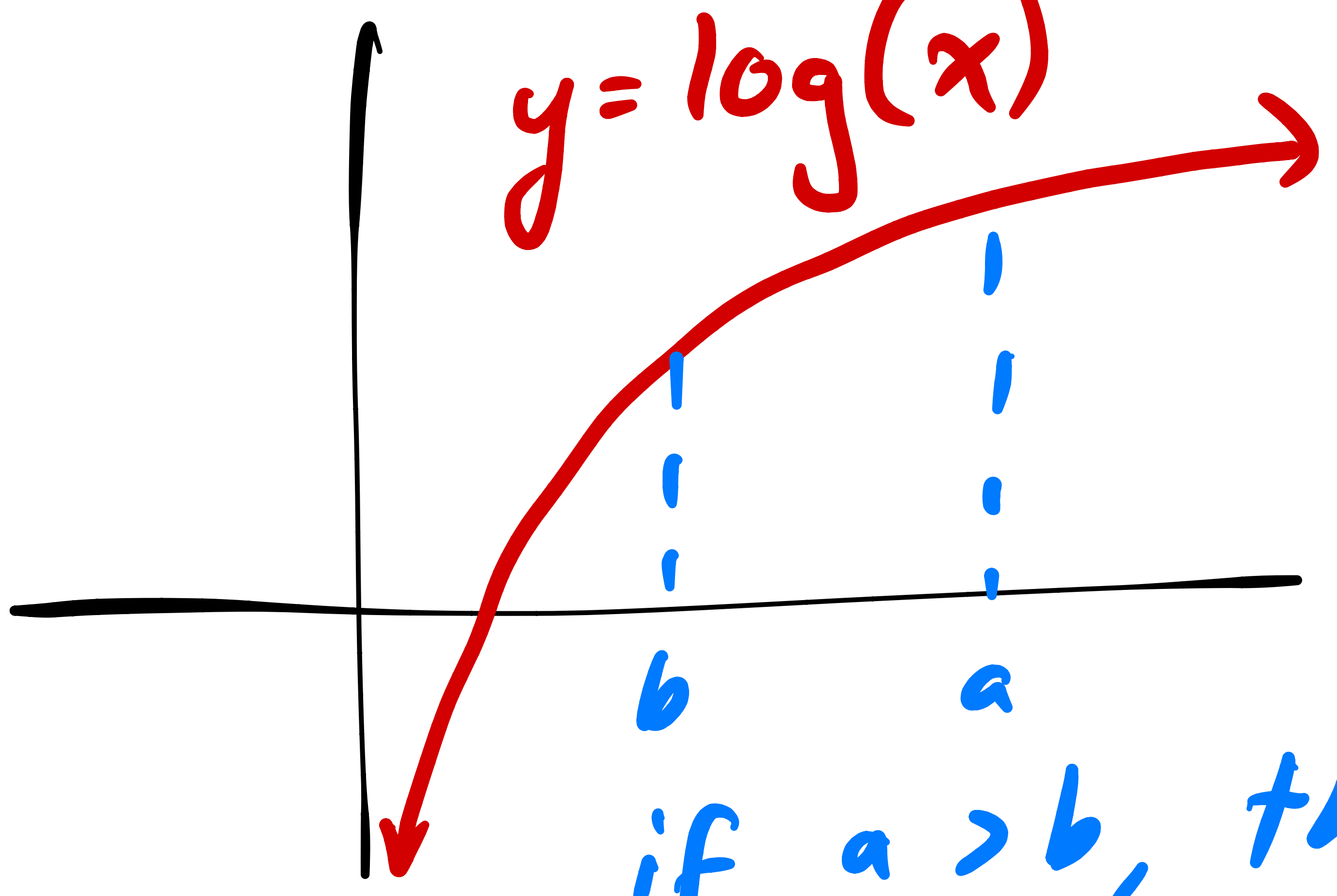
what if $p = 0.6 \rightarrow p(\text{sequence}) = 0.6 \cdot 0.4 \cdot 0.4 \cdot 0.6 \cdot 0.4^5 \cdot 0.6$

$= 0.6^3 \, 0.4^7$

what if $p = 0.2 \rightarrow p(\text{sequence}) = 0.2^3 \, 0.8^7$

$L(p) = P(\text{sequence given } p) = p^3 (1-p)^7 \rightarrow$ goal: maximize $L(p)$!

idea: maximize log-likelihood instead!!!



if $a > b$, then $\log(a) > \log(b)$

$$LL(p) = \log(L(p)) = \log\left(p^3(1-p)^7\right)$$

$$= \log(p^3) + \log\left((1-p)^7\right)$$

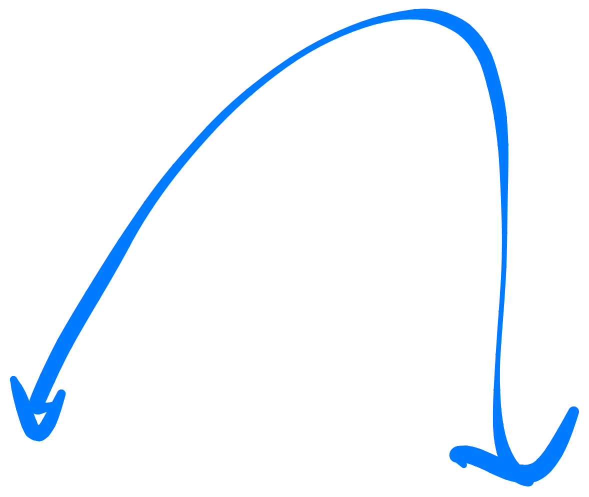
$$LL(p) = 3\log(p) + 7\log(1-p)$$

$$\frac{d}{dp} LL(p) = 3 \cdot \frac{1}{p} + 7 \cdot \frac{1}{1-p}(-1) = 0$$

$$\frac{3}{p} = \frac{7}{1-p}$$

$$3(1-p) = 7p$$

$$3 = 10p \Rightarrow \boxed{p = \frac{3}{10}}$$



log rules

$$\log(ab) = \log(a) + \log(b)$$

$$\log(a^n) = n\log(a)$$

Maximum likelihood estimation (MLE)

- In general, if we flip a fair coin n times and see x heads, to find the "best guess" for p, we **maximize** the **likelihood function**

$$L(p) = p^x(1-p)^{n-x}$$

- It is hard to maximize this directly, so instead we maximize the **log-likelihood:**

$$LL(p) = x \log p + (n-x)\ \log(1-p)$$

- This is maximized when $p = \dfrac{x}{n}$, which matches our intuitive guess.

MLE and regression

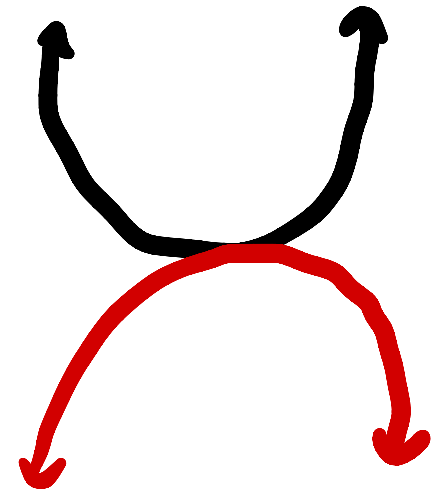
$$LL(a, b, \sigma^2) = -\frac{1}{n} \sum_{i=1}^{n} (y_i - a - b x_i)^2$$

$$y_i = a + b x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

normal distribution

- **Key point:** if you use the assumption that the errors in a linear regression model are independent and follow a Normal distribution, then **maximizing (log-)likelihood** is equivalent to minimizing **mean squared error**.



  - This is a big reason why least squares is so prevalent – not only is it computationally easy to minimize mean squared error, but it is consistent with this **assumption of normality**.
  - You're not expected to fully grasp this concept just yet, but it is valuable context to have throughout the rest of today's lecture, for tomorrow's lecture, and for the rest of your data science career.

Statistics in biology and sociology

Quetelet

- Adolphe Quetelet (1796-1874)[1] was a Belgian astronomer, mathematician, statistician, and sociologist.
    - He was born in Ghent (picture to the right).
- Originally an **astronomer**, he is known for being one of the first people to apply statistical methods to ideas in the **social sciences**.

Photo taken in Ghent by Suraj

1. https://www.britannica.com/biography/Adolphe-Quetelet

From astronomy to social science

- Astronomers took the **average** of several observations to estimate the **true value** of some quantity, i.e. to reduce observational error.
    - e.g. measuring the speed of Saturn.
- Quetelet was the first to apply the average to data on humans and societies.
    - For instance, he obtained a dataset containing the chest circumferences of thousands of Scottish soldiers[1].
    - He computed the **average** of the chest circumferences of these soldiers, yielding ~39.75 inches.
    - **What does this mean?**

1. https://www.theatlantic.com/business/archive/2016/02/the-invention-of-the-normal-person/463365/

The "average man"

- Possible interpretations:
  - 39.75 inches is the chest size we'd expect if we selected a random soldier.
  - 39.75 inches is roughly the chest size of a normal soldier.
- Quetelet's interpretation: 39.75 inches is the **true** chest size of soldiers, and any differences in an individual's chest size is due to error.
  - Later, Quetelet described the concept of the "average man" (in his words, *l'homme moyen*), who is defined by having an average measurement in several biological and sociological characteristics.
  - He believed that with more data, we could get closer and closer to approximating the "true" human.

Quetelet index

- To quantify the weight and height of the average man, Quetelet devised the **Quetelet index**, defined as

$$\text{Quetelet index} = \frac{\text{mass in kg}}{(\text{height in m})^2}$$

= Body Mass Index (BMI)

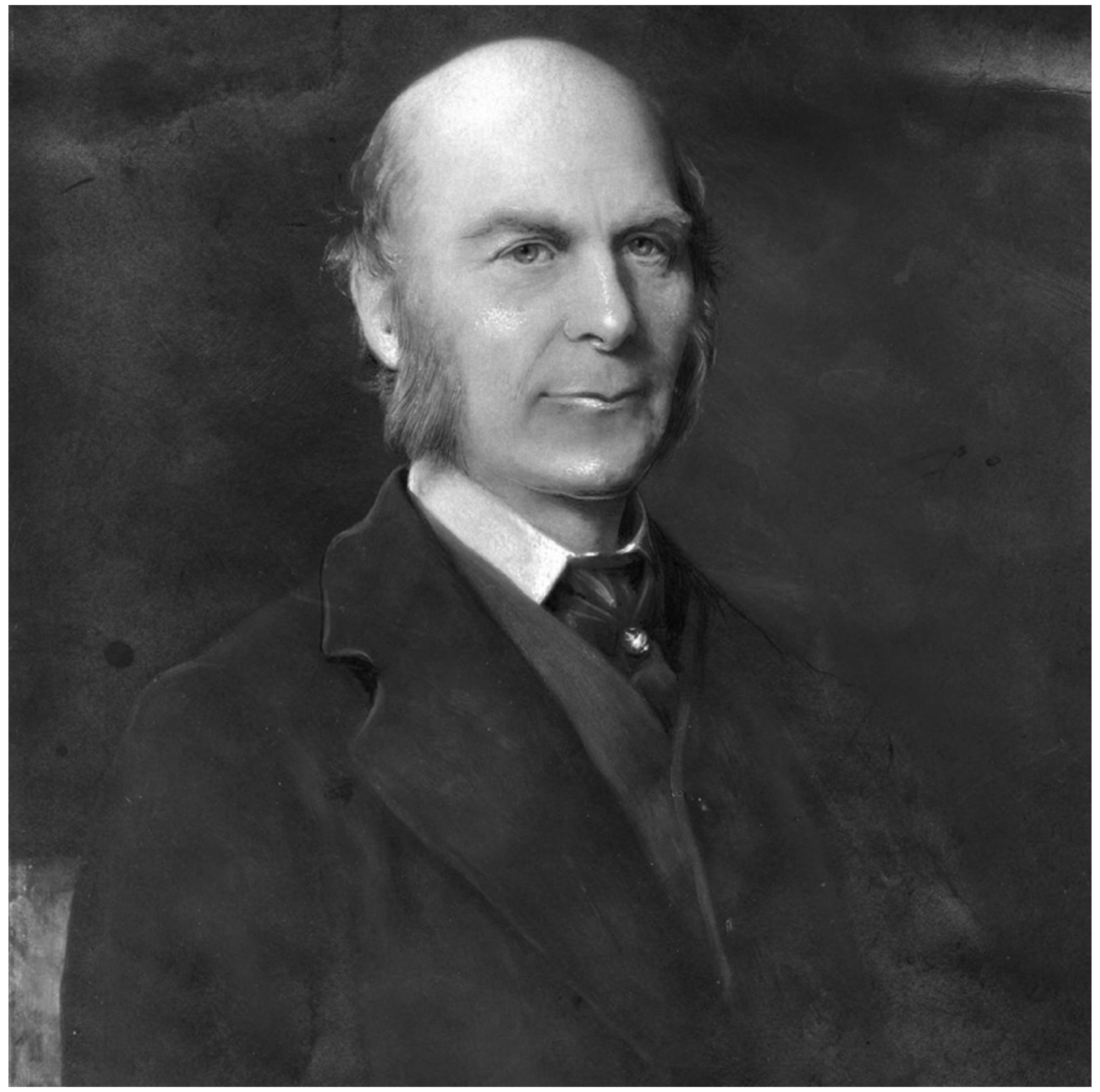
- Quetelet never intended it to be a measure of obesity.
- However, in the 1900s, the Quetelet index began to be used for this purpose by insurance companies.

Galton

Galton

Legrendre : 1805
Gauss : 1809

- Sir Francis Galton (1822-1911) was a British polymath.
  - He was knighted in 1909, hence the "Sir".
- He was Charles Darwin's half-cousin. As we will see, this played a pivotal role in the ideas he decided to study.
  - Example: Galton was the first to discover that everyone has unique fingerprints, and thus that fingerprints can be used for identification.
- galton.org contains excerpts of many of his original works.
- Note that Galton was only born years after Legendre and Gauss formulated least squares.



1. https://www.britannica.com/biography/Francis-Galton

Aside: Darwin and the "Rule of Three"

$$\frac{a}{b} = \frac{c}{???}$$

- Charles Darwin is well-known for his development of the **theory of evolution**, though he was supposedly not found of mathematics.
- He wrote to a colleague,

  > *"I have no faith in anything short of actual measurement and the Rule of Three."*[1]

- The "Rule of Three" in question is that if $\frac{a}{b} = \frac{c}{d}$, then given any three of a, b, c, d, one can find the fourth by cross-multiplication.
- **Question:** does the Rule of Three work when there is measurement error in any of a, b, c, d?

1. Stigler, *The Seven Pillars of Statistical Wisdom*, p.107

$$\frac{10 \text{ miles}}{20 \text{ min}} = \frac{15 \text{ miles}}{???}$$

what if

$$\frac{10.5}{19} = \frac{15}{???}$$

→ rule of three doesn't work when there's measurement error

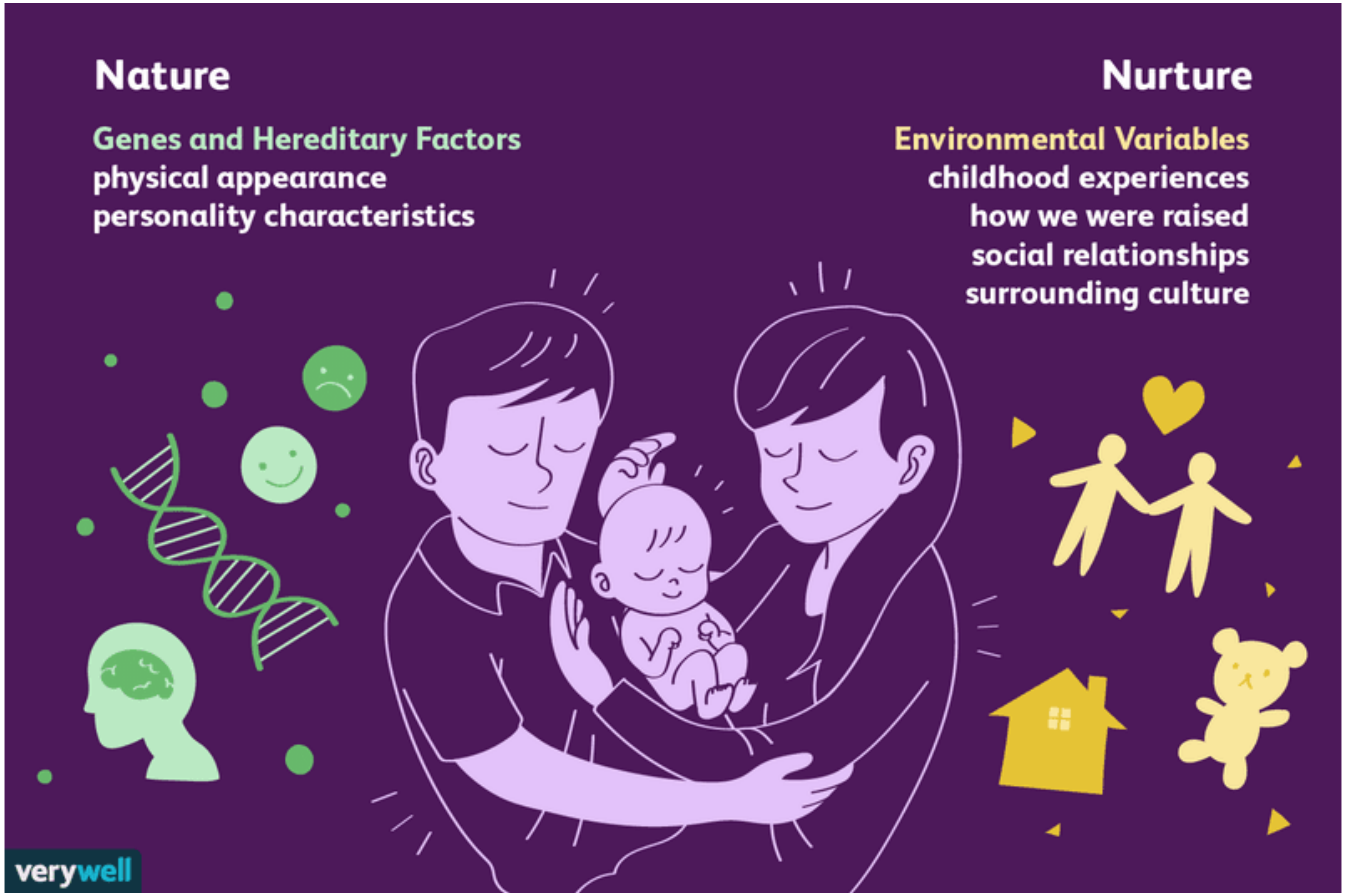
Galton's motivation

- Galton was interested in studying how traits were passed from parents to children.
- He created the field of **eugenics**, and wrote:

  "Eugenics is the science which deals with all influences that improve the inborn qualities of a race; also with those that develop them to the utmost advantage. The improvement of the inborn qualities, or stock, of some one human population, will alone be discussed here."[1]

  - In other words, he believed that the traits that made people successful were inheritable, and so only people with those characteristics should have children.
  - Along these lines, he "ranked" the worth of each race.
- Virtually all of the statistical techniques he developed were to further his study of eugenics.

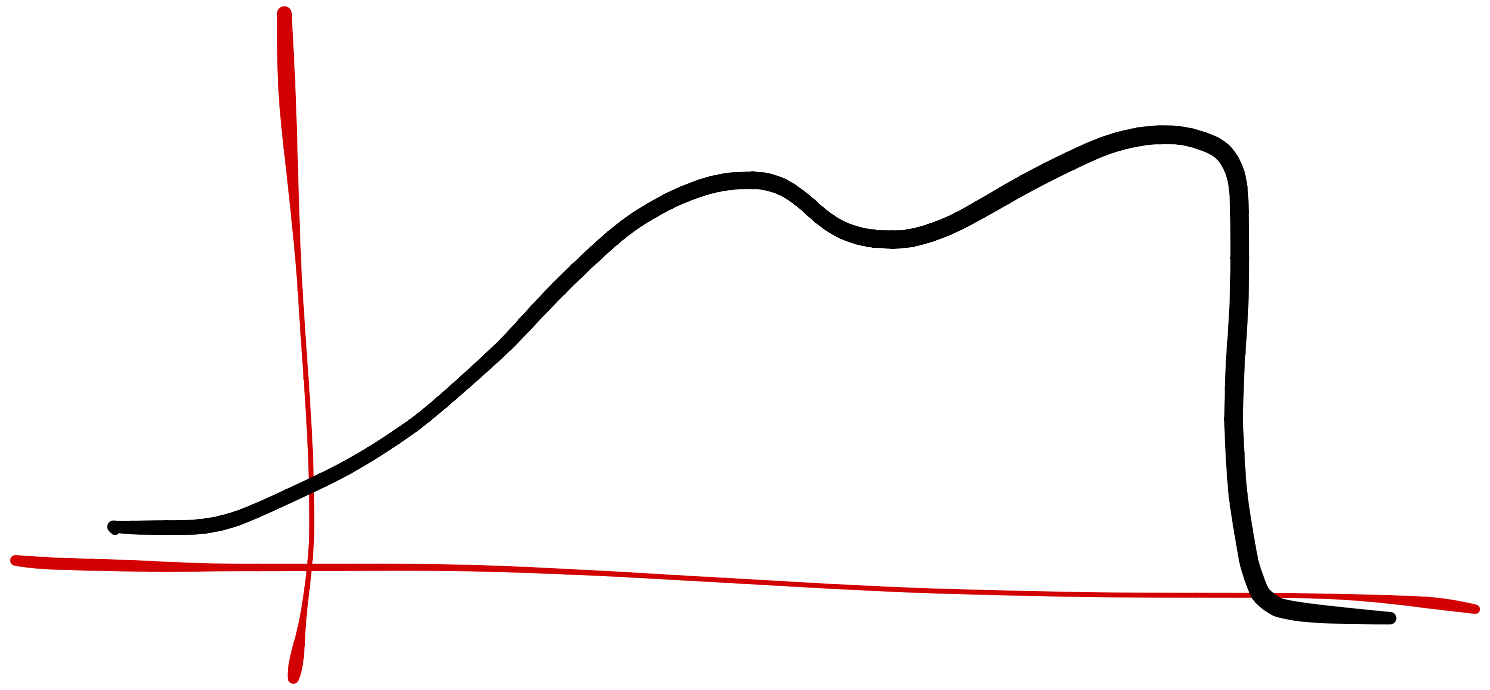
1. https://galton.org/essays/1900-1911/galton-1905-socpapers-eugenics-definition-scope-aims.pdf



Galton coined the phrase "nature vs. nurture." (graphic source)

Percentiles

p^{th} percentile = the smallest value greater than or equal to p% of values



- Galton developed the idea of a **percentile**.
- He collected data on the physical measurements of many individuals and summarized the data using percentiles, quartiles, and deciles.

WHEN any large group of statistical cases is sorted into a hundred classes equal in number, and progressively increasing in value, the dividing values between the classes are called Percentiles;[2] or, if into ten classes, they are called Deciles; or if into four classes, they are called Quartiles. The fiftieth percentile, the fifth decile, and the second quartile are consequently the same as the median. All other deciles, &c., are calculated on the

etc = et cetera

1. https://galton.org/essays/1880-1889/galton-1885-nature-percentiles.pdf

(*The value that is unreached by n per cent. of any large group of measurements, and surpassed by 100−n of them, is called its n*th *percentile*)

percentiles

Subject of measurement	Age	Unit of measurement	Sex	No. of persons in the group	Values surpassed by per-cents as below: 95 / Values unreached by per-cents as below: 5	90 / 10	80 / 20	70 / 30	60 / 40	50 / 50	40 / 60	30 / 70	20 / 80	10 / 90	5 / 95
Height, standing, without shoes ..	23-51	Inches	M.	811	63·2	64·5	65·8	66·5	67·3	67·9	68·5	69·2	70·0	71·3	72·4
			F.	770	58·8	59·9	61·3	62·1	62·7	63·3	63·9	64·6	65·3	66·4	67·3
Height, sitting, from seat of chair ...	23-51	Inches	M.	1013	33·6	34·2	34·9	35·3	35·4	36·0	36·3	36·7	37·1	37·7	38·2
			F.	775	31·8	32·3	32·9	33·3	33·6	33·9	34·2	34·6	34·9	35·6	36·0
Span of arms ...	23-51	Inches	M.	811	65·0	66·1	67·2	68·2	69·0	69·9	70·6	71·4	72·3	73·6	74·8
			F.	770	58·6	59·5	60·7	61·7	62·4	63·0	63·7	64·5	65·4	66·7	68·0
Weight in ordinary indoor clothes ..	23-26	Pounds	M.	520	121	125	131	135	139	143	147	150	156	165	172
			F.	276	102	105	110	114	118	122	129	132	136	142	149
Breathing capacity	23-26	Cubic inches	M.	212	161	177	187	199	211	219	226	236	248	277	290
			F.	277	92	102	115	124	131	138	144	151	164	177	186
Strength of pull as archer with bow	23-26	Pounds	M.	519	56	60	64	68	71	74	77	88	82	89	96
			F.	276	30	32	34	36	38	40	42	44	47	51	54
Strength of squeeze with strongest hand	23-26	Pounds	M.	519	67	71	76	79	82	85	88	91	95	100	104
			F.	276	36	39	43	47	49	52	55	58	62	67	72
Swiftness of blow.	23-26	Feet per second	M.	516	13·2	14·1	15·2	16·2	17·3	18·1	19·1	20·0	20·9	22·3	23·6
			F.	271	9·2	10·1	11·3	12·1	12·8	13·4	14·0	14·5	15·1	16·3	16·9
Sight, keenness of—by distance of reading diamond test-type	23-26	Inches	M.	398	13	17	20	22	23	25	26	28	30	32	34
			F.	433	10	12	16	19	22	24	26	27	29	31	32

1. https://galton.org/essays/1880-1889/galton-1885-nature-percentiles.pdf

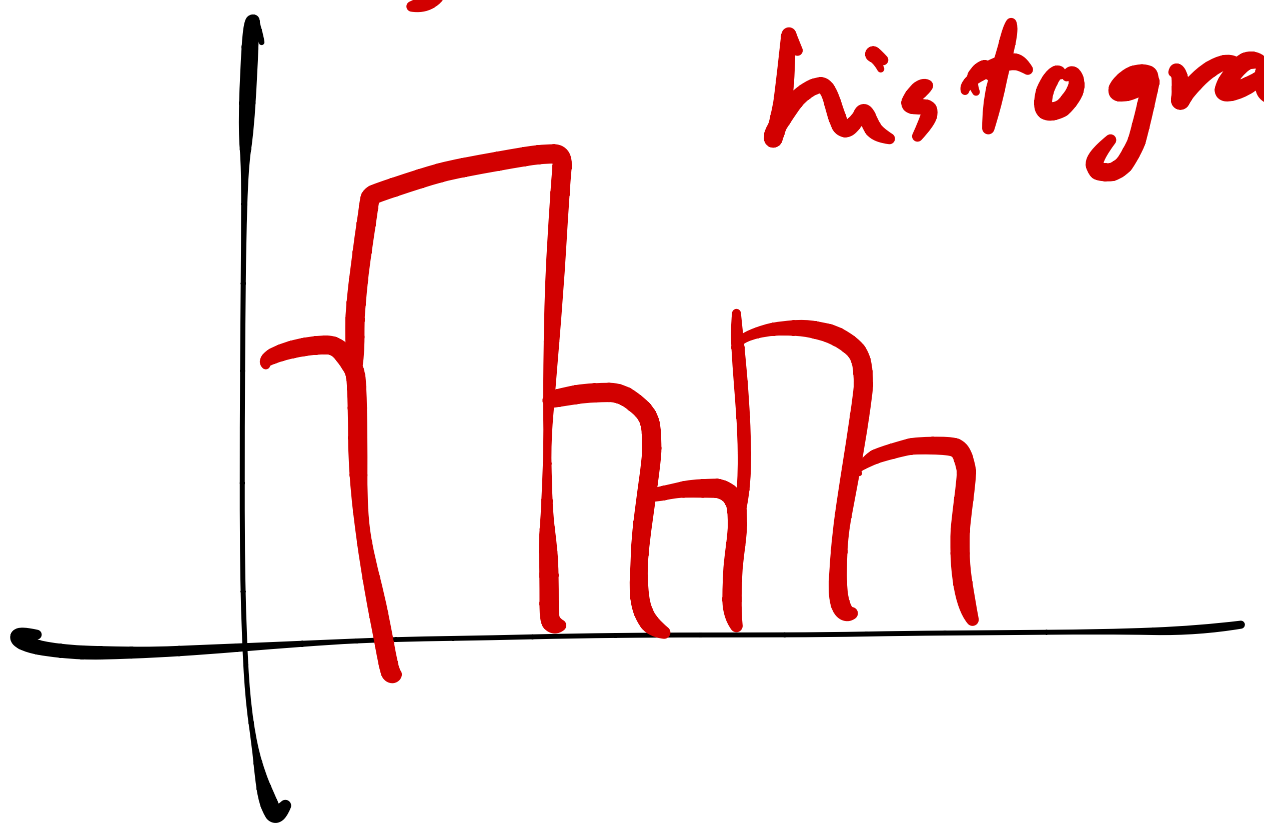
poverty

Table A.

(A.) Pauperism. Per Cent.	(B.) No. of Unions.	(C.) Sums of B from top.	(D.) Successive tenths of the total of B.	(E.) D – C (in each row).	(F.) D – C multiplied into 0·5.	(G.) D – C × (0·5) and divided by B_1 *	Interpolated Deciles Order.	Value (G + A) †
Below 1·75	7	7	—	—	—	—	—	—
1·75 to 2·25	7	14	—	—	—	—	—	—
2·25 ,, 2·75	11	25	—	—	—	—	—	—
2·75 ,, 3·25	21	46	59	13	6·5	0·23	1st	3·48
3·25 ,, 3·75	28	74	—	—	—	—	—	—
3·75 ,, 4·25	33	107	118	11	5·5	0·12	2nd	4·37
4·25 ,, 4·75	46	153	176	23	11·5	0·21	3rd	4·96
4·75 ,, 5·25	55	208	235	27	13·5	0·34	4th	5·59
5·25 ,, 5·75	40	248	—	—	—	—	—	—
5·75 ,, 6·25	45	293	294	1	0·5	0·01	5th	6·26
6·25 ,, 6·75	44	337	353	16	8·0	0·23	6th	6·98
6·75 ,, 7·25	35	372	412	40	20 0	0·45	7th	7·70
7·25 ,, 7·75	44	416	—	—	—	—	—	—
7·75 ,, 8·25	31	447	470	23	11·5	0·43	8th	8·68
8·25 ,, 8 75	27	474	—	—	—	—	—	—
8·75 ,, 9·25	34	508	—	—	—	—	—	—
9·25 ,, 9·75	21	529	529	0	0·0	0·00	9th	9·75
9·75 ,, 10·25	11	540	—	—	—	—	—	—
Above 10·25	48	588	—	—	—	—	—	—
	588	—	—	—	—	—	—	—

* B_1 in column G means the entry in column B that lies *one line below* that on which the entry in F is standing. Thus 6·5 is divided by 28, and 5·5 by 46.

† The second decimal is approximate.



Here is an example of how Galton applied his **method of deciles**. Let's see if we can understand how it works.

1. https://galton.org/essays/1890-1899/galton-1896-jrss-percentiles-yule.pdf

Table A.

(A.) Pauperism. Per Cent.	(B.) No. of Unions.	(C.) Sums of B from top.	(D.) Successive tenths of the total of B.	(E.) D − C (in each row).	(F.) D − C multiplied into 0·5.	(G.) D − C × (0·5) and divided by B_1 *	Interpolated Deciles: Order.	Interpolated Deciles: Value (G + A) †
Below 1·75	7	7	—	—	—	—	—	—
1·75 to 2·25	7	14	—	—	—	—	—	—
2·25 „ 2·75	11	25	—	—	—	—	—	—
2·75 „ 3·25	21	46	59	13	6·5	0·23	1st	3·48
3·25 „ 3·75	28	74	—	—	—	—	—	—
3·75 „ 4·25	33	107	118	11	5·5	0·12	2nd	4·37
4·25 „ 4·75	46	153	176	23	11·5	0·21	3rd	4·96
4·75 „ 5·25	55	208	235	27	13·5	0·34	4th	5·59
5·25 „ 5·75	40	248	—	—	—	—	—	—
5·75 „ 6·25	45	293	294	1	0·5	0·01	5th	6·26
6·25 „ 6·75	44	337	353	16	8·0	0·23	6th	6·98
6·75 „ 7·25	35	372	412	40	20 0	0·45	7th	7·70
7·25 „ 7·75	44	416	—	—	—	—	—	—
7·75 „ 8·25	31	447	470	23	11·5	0·43	8th	8·68
8·25 „ 8 75	27	474	—	—	—	—	—	—
8·75 „ 9·25	34	508	—	—	—	—	—	—
9·25 „ 9·75	21	529	529	0	0·0	0·00	9th	9·75
9·75 „ 10·25	11	540	—	—	—	—	—	—
Above 10·25	48	588	—	—	—	—	—	—
	588	—	—	—	—	—	—	—

* B_1 in column G means the entry in column B that lies *one line below* that on which the entry in F is standing. Thus 6·5 is divided by 28, and 5·5 by 46.

† The second decimal is approximate.

height	count	cumulative counts
52-56	4	4
56-60	8	12 15
60-64	12	24
64-68	15	39
68-72	12	51
72-76	8	59
76-80	1	60
	60	

Suppose we want to find quartiles.

60 numbers total

→ # at pos 15,
at pos 30,
at pos 45

→ 60" has 12 values LEQ

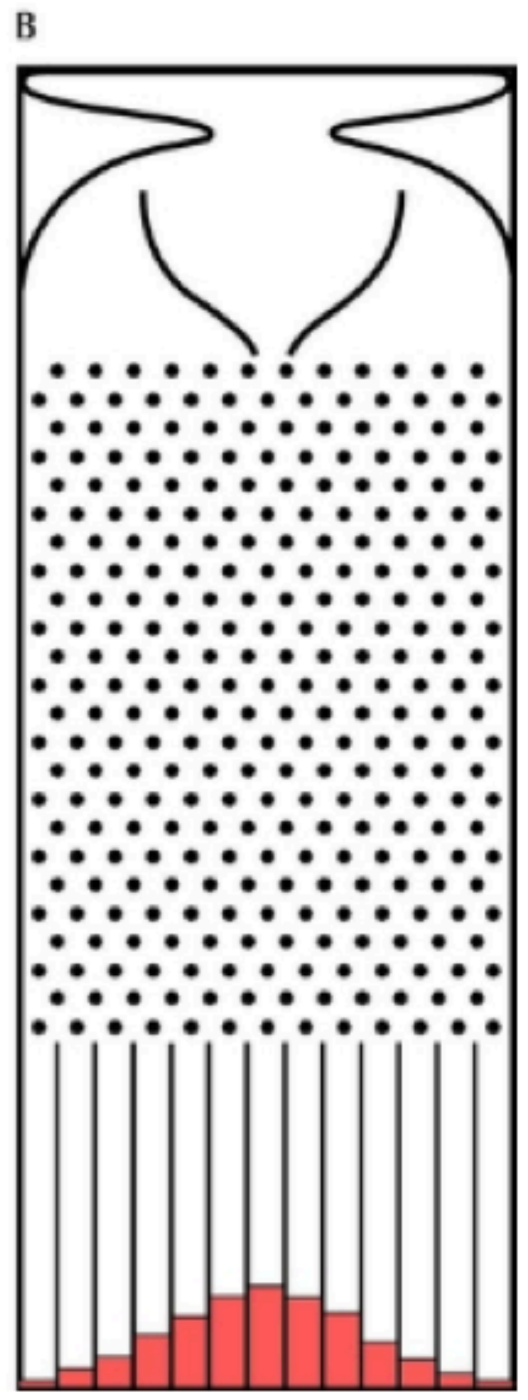
⇒ 64" has 24 values LEQ

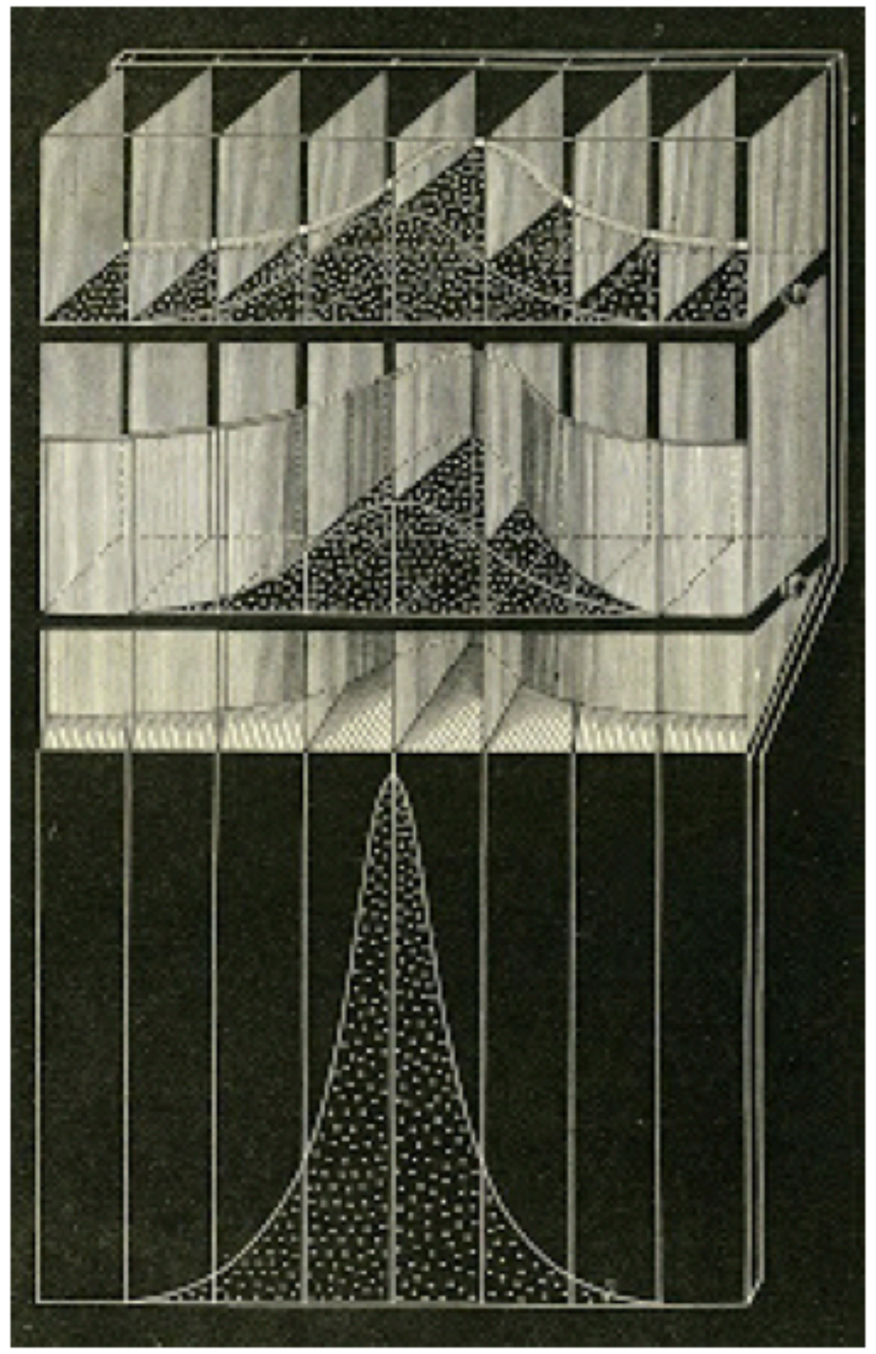
⇒ # at pos 15 is between 60" and 64"

Galton: $60 + 4 \cdot \left(\frac{15-12}{12}\right) = 60 + 4 \cdot \frac{1}{4} = 61$

Human characteristics are "normally" distributed

- While he did not derive the normal distribution (Gauss did), Galton was the first to call it by the name "normal" distribution, rather than the name "error distribution".
  - He observed that many human characteristics, such as human height, roughly follow a **normal** distribution.
  - In order to demonstrate why this is the case, he constructed what is known as a **quincrux**.
- See an animated quincrux here.





Heights

correlation coefficient (r) = the mean of the product of x and y, when both are measured in SU
↗ standard units

- One trait Galton was interested in studying was the difference in heights between parents and their children.
- He defined a new quantity, "midparent height", as being the average of a child's mother's and father's heights, after the mother's height was multiplied by 1.08.
  - He also multiplied the heights of daughters by 1.08.
- After collecting data, he estimated that the correlation between the **deviations of midparent heights** and the **deviations of child heights** was $\frac{2}{3}$.

TABLE I.

Number of Adult Children of various statures born of 205 Mid-parents of various statures.

(All Female heights have been multiplied by 1·08).

Heights of the Mid-parents in inches.	Heights of the Adult Children.														Total Number of		Medians.
	Below	62·2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above	Adult Children.	Mid-parents.	
Above ..	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	..
72·5	..	..	..	..	..	..	..	1	2	1	2	7	2	4	19	6	72·2
71·5	..	..	..	..	1	3	4	3	5	10	4	9	2	2	43	11	69·9
70·5	1	..	1	..	1	1	3	12	18	14	7	4	3	3	68	22	69·5
69·5	..	..	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68·9
68·5	1	..	7	11	16	25	31	34	48	21	18	4	3	..	219	49	68·2
67·5	..	3	5	14	15	36	38	28	38	19	11	4	..	..	211	33	67·6
66·5	..	3	3	5	2	17	17	14	13	4	..	..	..	..	78	20	67·2
65·5	1	..	9	5	7	11	11	7	7	5	2	1	..	..	66	12	66·7
64·5	1	1	4	4	1	5	5	..	2	..	..	..	..	..	23	5	65·8
Below ..	1	..	2	4	1	2	2	1	1	..	..	..	..	..	14	1	..
Totals ..	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	..
Medians ..	..	..	66·3	67·8	67·9	67·7	67·9	68·3	68·5	69·0	69·0	70·0	..	..	..	..	..

Note.—In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62·2, 63·2, &c., instead of 62·5, 63·5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents.

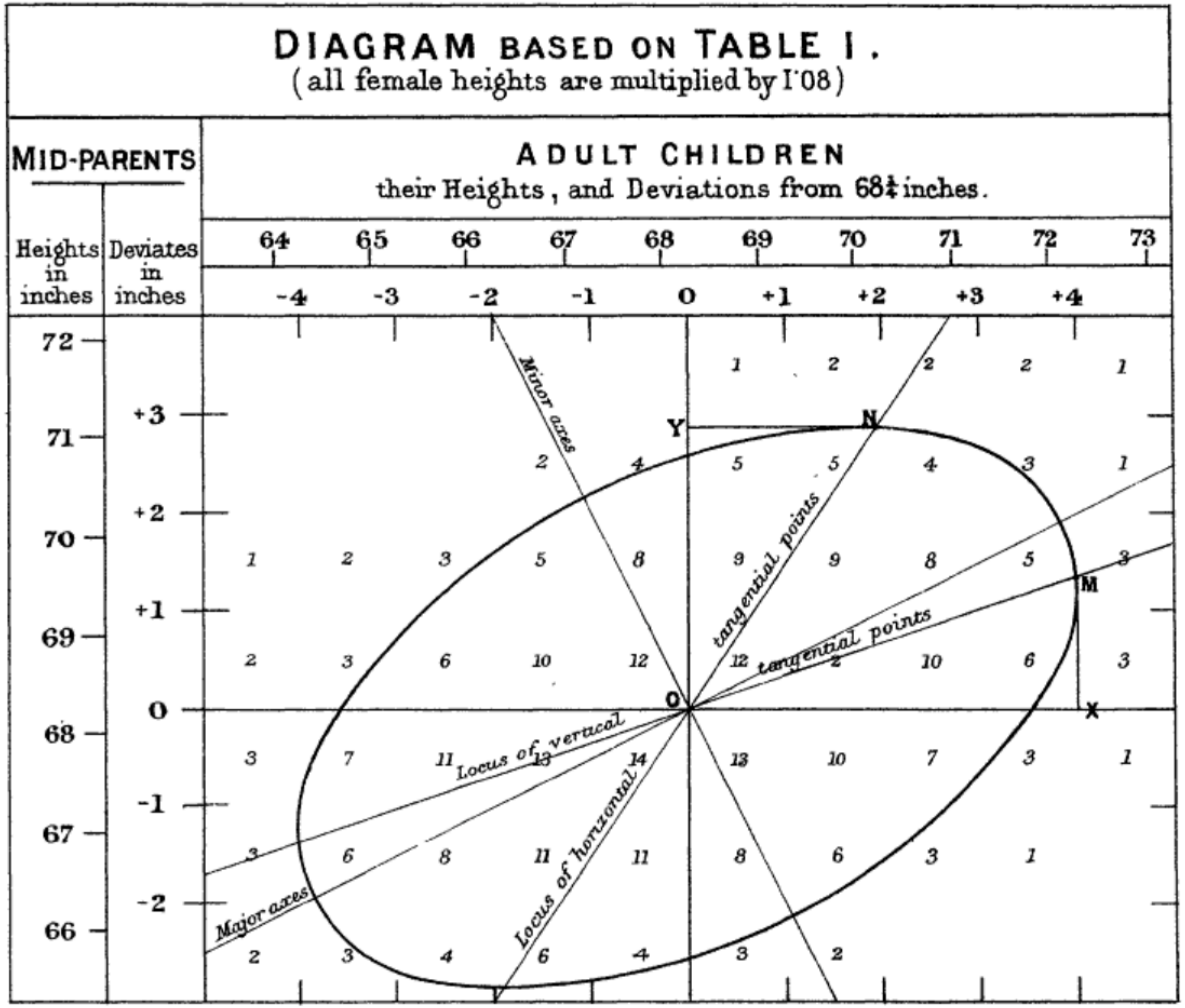
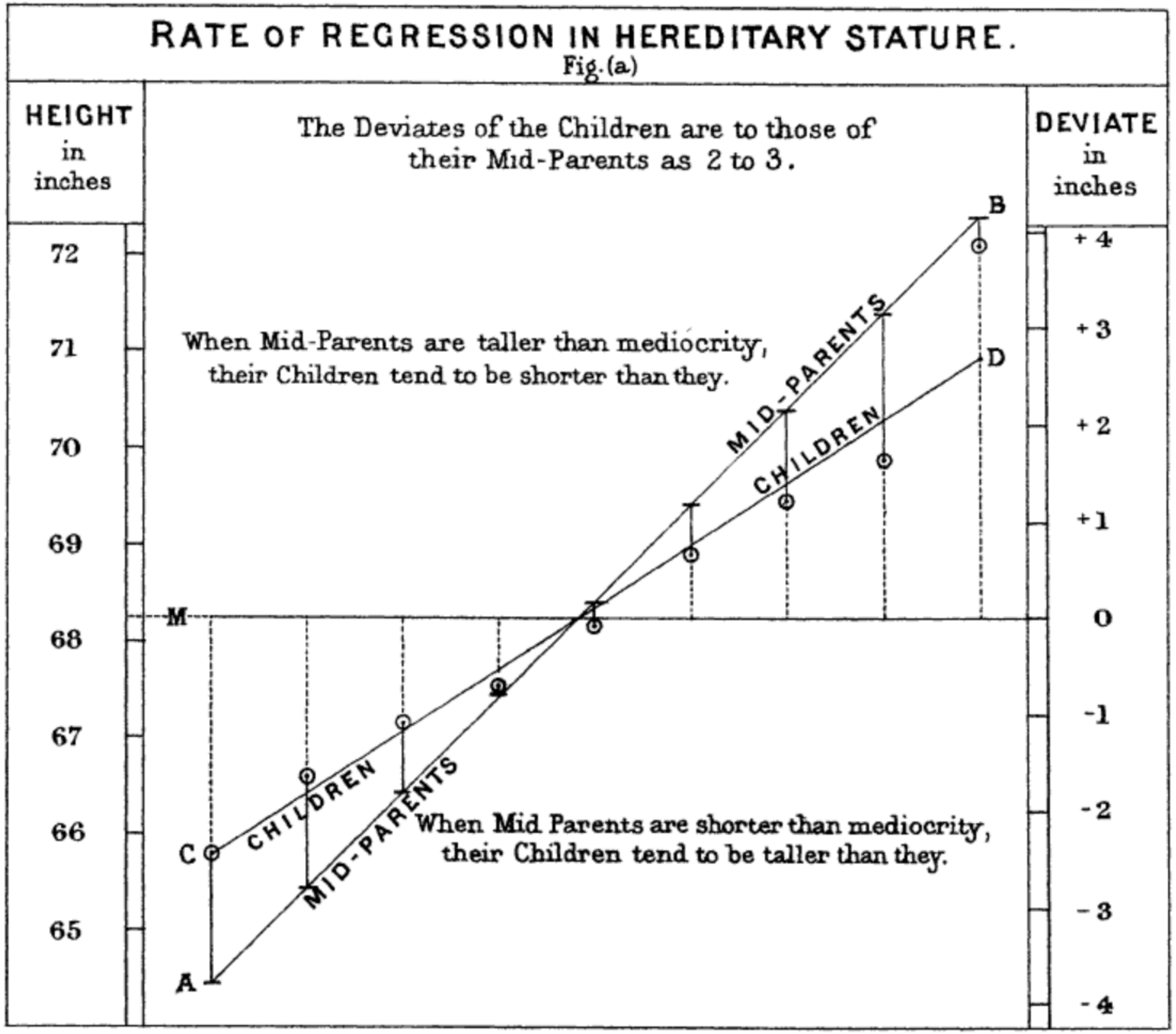


DIAGRAM BASED ON TABLE I.
(all female heights are multiplied by 1·08)
MID-PARENTS
Heights in inches
Deviates in inches
ADULT CHILDREN
their Heights, and Deviations from 68¼ inches.
64 65 66 67 68 69 70 71 72 73
-4 -3 -2 -1 0 +1 +2 +3 +4
72 71 70 69 68 67 66
+3 +2 +1 0 -1 -2
Minor axes
Major axes
tangential points
tangential points
Locus of vertical
Locus of horizontal
Y
N
M
X
O


RATE OF REGRESSION IN HEREDITARY STATURE.
Fig. (a)
HEIGHT in inches
DEVIATE in inches
The Deviates of the Children are to those of their Mid-Parents as 2 to 3.
When Mid-Parents are taller than mediocrity, their Children tend to be shorter than they.
When Mid Parents are shorter than mediocrity, their Children tend to be taller than they.
MID-PARENTS
CHILDREN
MID-PARENTS
CHILDREN
A
B
C
D
M
72
71
70
69
68
67
66
65
+4
+3
+2
+1
0
-1
-2
-3
-4

and breadth (0·45). The concluding passage of the memoir is worth citing for its historical interest:—"The prominent "characteristics of any two correlated variables, so far at least as "I have as yet tested them, are four in number. It is supposed "that their respective measures have been first transmuted into "others of which the unit is in each case equal to the probable "error of a single measure in its own series. Let y = the deviation "of the subject, whichever of the two variables may be taken in "that capacity; and let x_1, x_2, x_3, &c., be the corresponding "deviations of the relative, and let the mean of these be X. Then "we find (1) that $y = r\text{X}$ for all values of y, (2) that r is the same "whichever of the two variables is taken for the subject, (3) that r "is always less than 1, (4) that r measures the closeness of the "co-relation." Galton determined r by a simple graphic method, Table of Data for calculating Tables of Distribution of Stature among the Kinsmen of Persons whose Stature is known.

From group of persons of the same Stature, to their Kinsmen in various near degrees.	Mean regression $= w$.	$Q = f$ $= p \times \sqrt{(1 - w^2)}$.
Mid-parents to Sons	2 / 3	1·27
Brothers to Brothers	2 / 3	1·27
Fathers or Sons to Sons or Fathers	1 / 3	1·60
Uncles or Nephews to Nephews or Uncles	2 / 9	1·66
Grandsons to Grandparents...	1 / 9	Practically that of Population, or 1·7 inch.
Cousins to Cousins	2 / 27	

Regression to the mean

- The effect that Galton observed was that **children tended to have heights that were closer to average than their parents.**
  - Tall parents tended to have children that were still tall, but closer to the average child's height.
  - Short parents tended to have children that were still short, but closer to the average child's height.
  - The same effect holds true in the opposite direction – remember, the correlation coefficient is symmetric!
- He called this "**reversion** to the mean", and later "**regression** to the mean".
- **The presence of regression to the mean depends on random variability in the distributions from which observations are drawn.**

Pearson

- Karl Pearson (1857-1936), a British statistician, was one of Galton's disciples.
  - He was also a stauch eugenicist.
  - He further developed the theory of correlation, and defined the correlation coefficient as we know it now.
- He founded the world's first Statistics department, at University College London, in 1911.
  - Started as part of UCL's Eugenics department.
  - Fun fact: UCSD has the world's first Cognitive Science department!

Summary, next time

Summary, next time

- Both Legendre and Gauss developed the theory of least squares, but Gauss tied it to probability theory.
    - Both used least squares in the development of planetary models.
- Quetelet was one of the first to apply tools from statistics to the social sciences, and was interested in studying the composition of the "average man".
- Galton pioneered many now-ubiquitous ideas in statistics, including that of the percentile and the term "regression".
    - He was motivated by the study of inheritance, and more specifically **eugenics**.