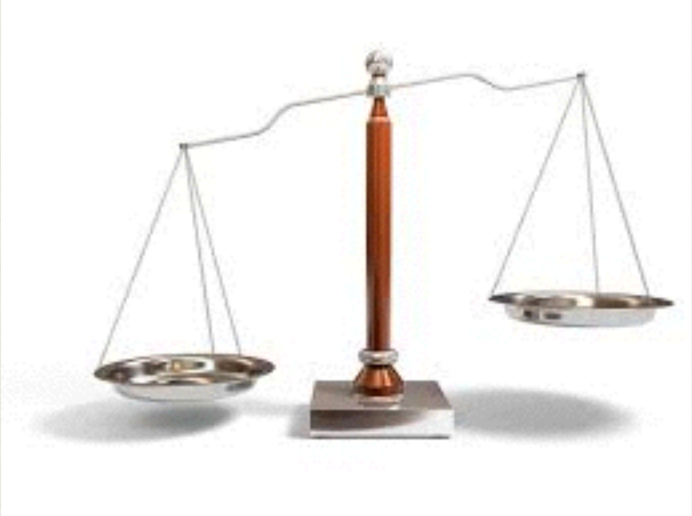
The Law of Errors



Suppose we measure the weight of a grain of rice.

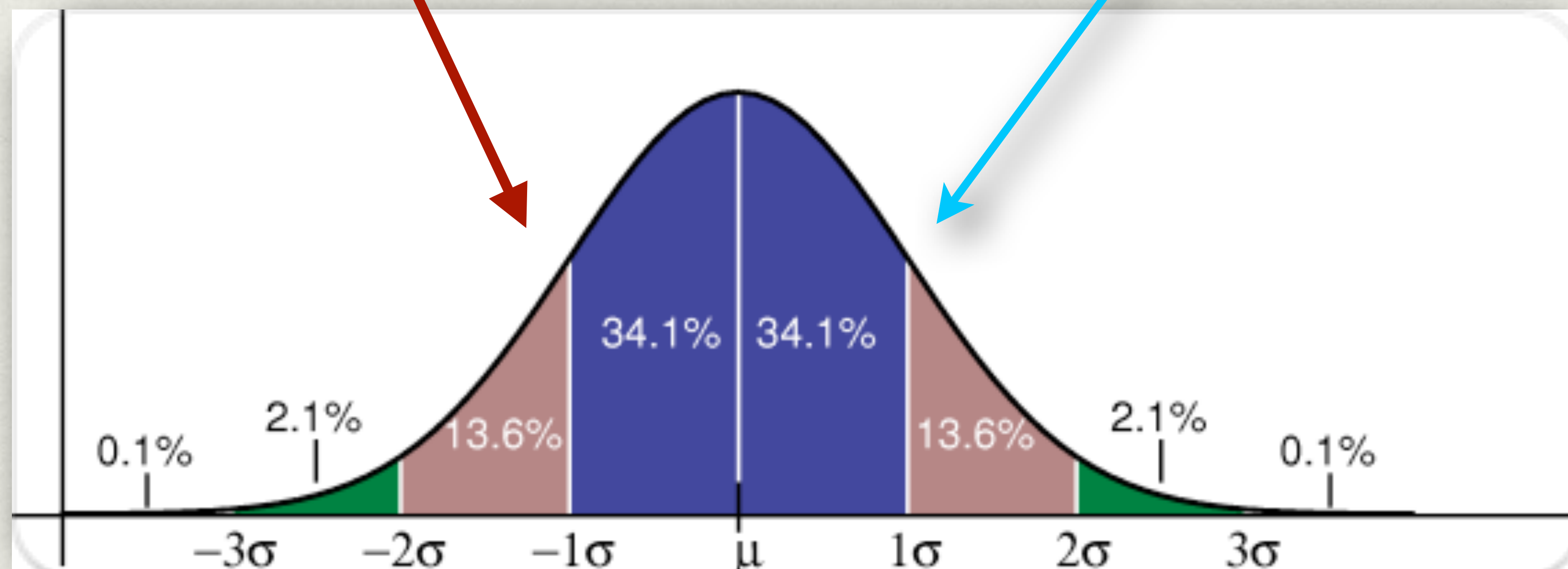
We take lost of measurements of the same rice grain.

What is the true or actual weight?

.025 grams .033 grams .029 grams

1 standard deviation below the average

1 standard deviation above the average



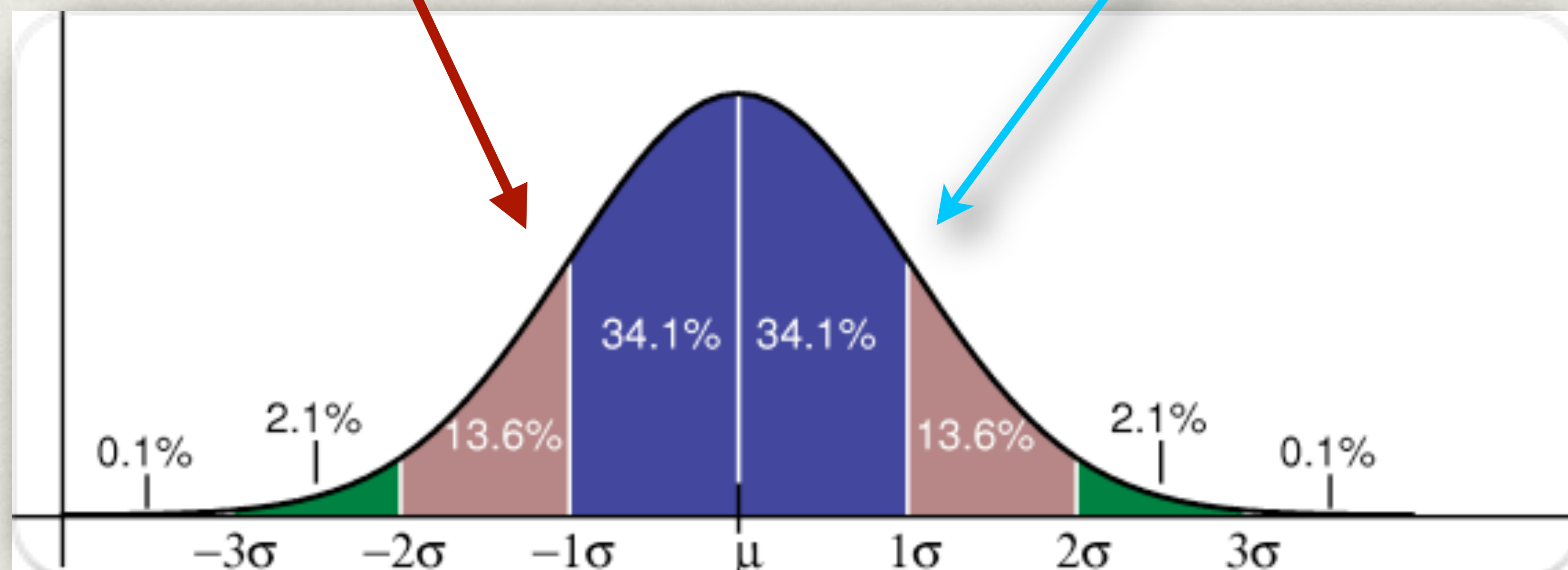
.028

An average grain of rice weighs around .028 grams

If we took 1000 measurements of the weight of the same grain of rice, our measurements would not always be the same.

1 standard deviation below the average

1 standard deviation above the average



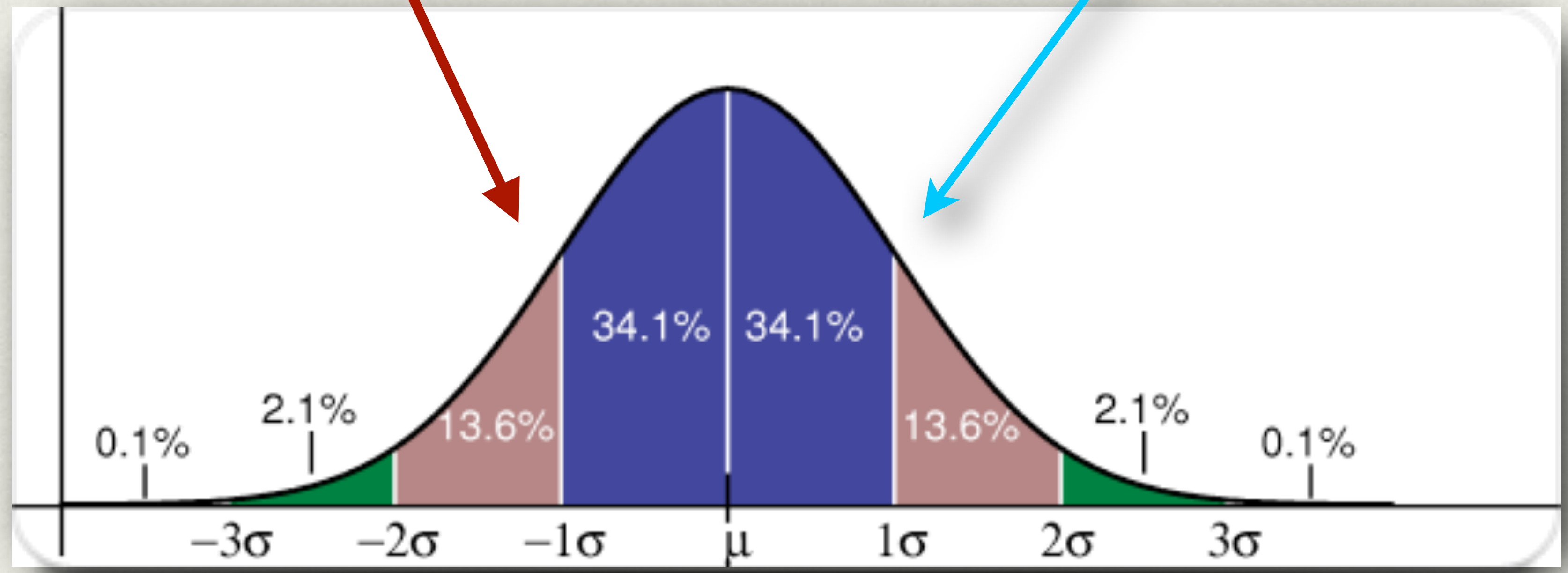
.028

An average grain of rice weighs around .028 grams

Why are our measurements not always the same?

1 standard deviation below the average

1 standard deviation above the average



.028

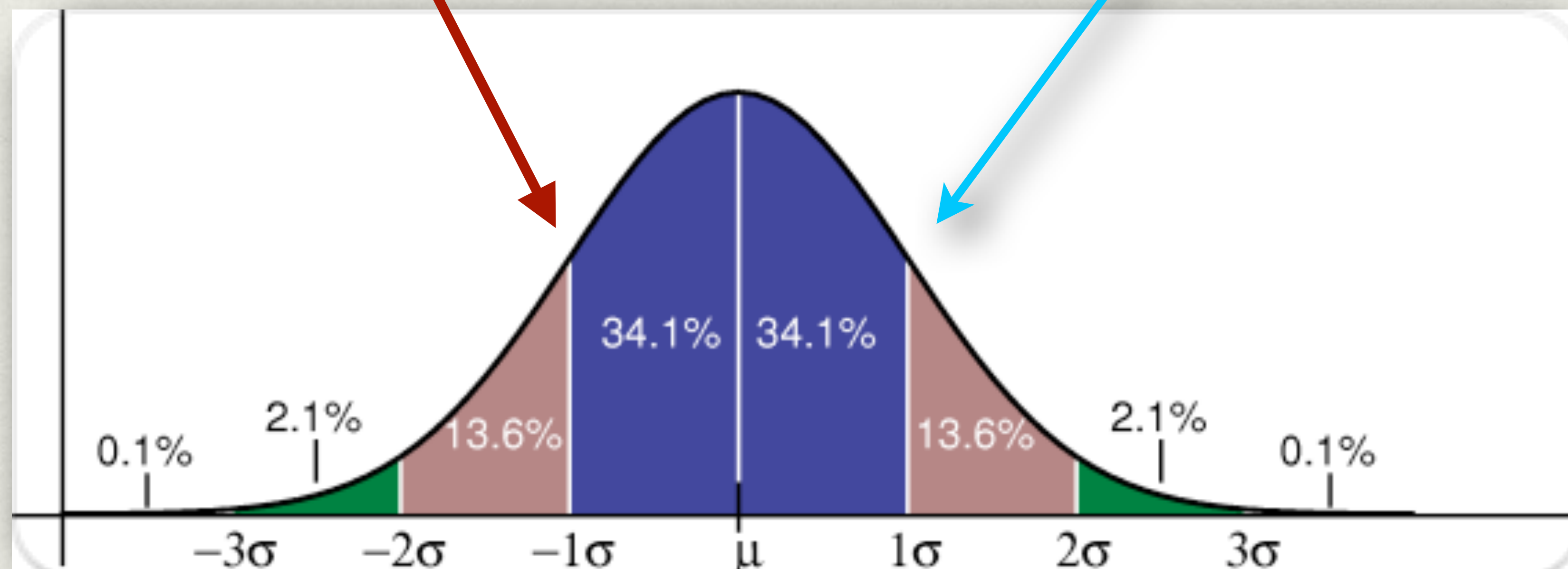
An average grain of rice weighs around .028 grams

<u>Why</u> are they not always the same?

Because the measurement instrument we use is itself never perfectly reliable. And we as observers are also not perfectly reliable.

34.1% of measurements are 1 standard deviation **below** the average

34.1% of measurements are 1 standard deviation **above** the average



.028

An average grain of rice weighs around .028 grams

The Law of Errors (in our measurements) is represented in the Gaussian (another term is "Normal") distribution.

The standard deviation

Consider a population consisting of the following eight values:

$$2,\ 4,\ 4,\ 4,\ 5,\ 5,\ 7,\ 9.$$

These eight data points have the mean (average) of 5:

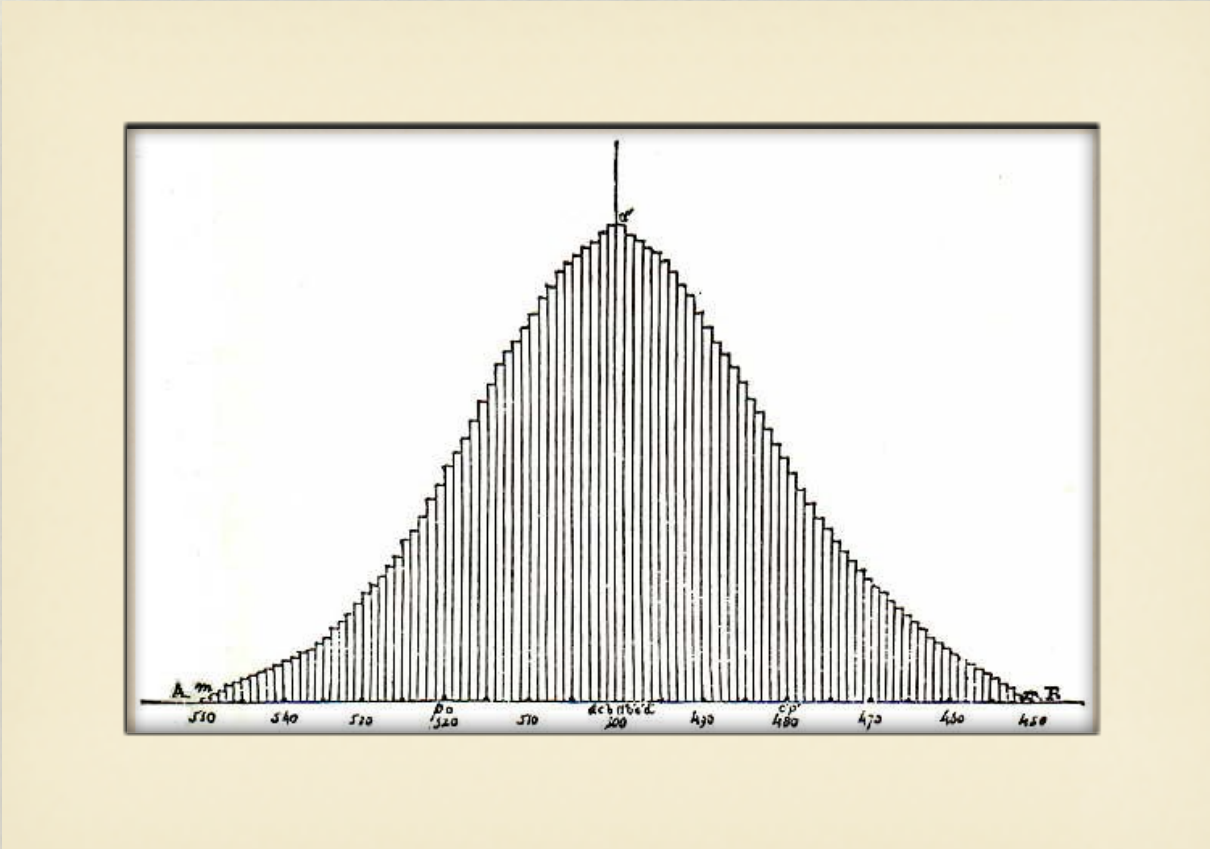
$$\frac{2+4+4+4+5+5+7+9}{8} = 5.$$

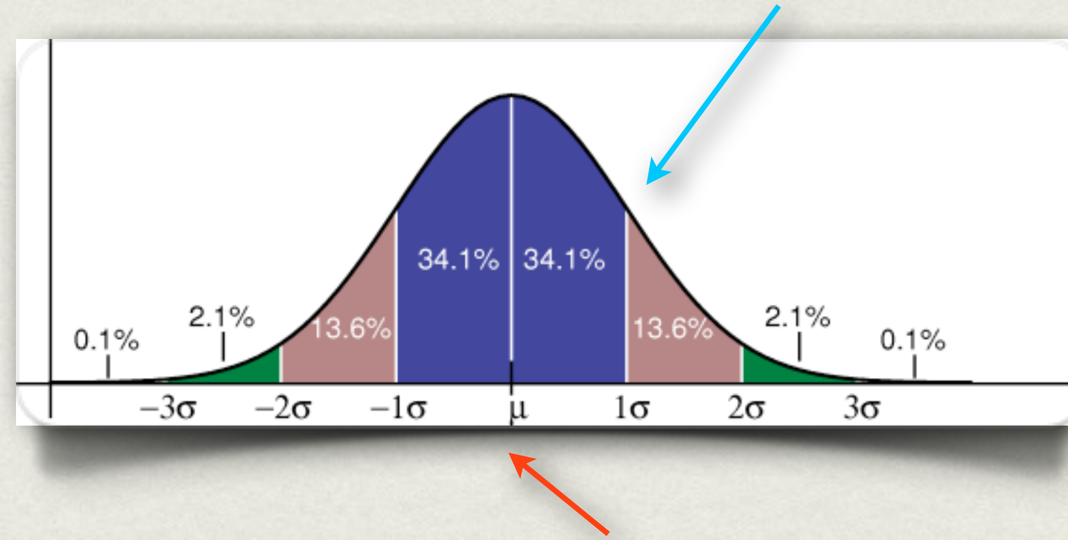
To calculate the population standard deviation, first compute the difference of each data point from the mean, and square the result of each:

$$\begin{array}{ll} (2-5)^2 = (-3)^2 = 9 & (5-5)^2 = 0^2 = 0 \\ (4-5)^2 = (-1)^2 = 1 & (5-5)^2 = 0^2 = 0 \\ (4-5)^2 = (-1)^2 = 1 & (7-5)^2 = 2^2 = 4 \\ (4-5)^2 = (-1)^2 = 1 & (9-5)^2 = 4^2 = 16. \end{array}$$

Next, compute the average of these values, and take the square root:

$$\sqrt{\frac{(9+1+1+1+0+0+4+16)}{8}} = 2.$$





A distribution of measurements recording the weight of a <u>**single**</u> grain of rice. We have taken a **thousand** measurements of **one** grain.

A distribution of the height of 1000 <u>**individual**</u> soldiers. We have taken **one** measurement of a **thousand** recruits.

This distribution of numbers represents **errors** in our measurements or the deviation from the true value.

What does this distribution of numbers represent?

Quetelet

Quetelet was struck by the fact that a plot of variation in the frequency of height around a population mean gave a result that conformed exactly to the bell-shaped curve predicted by the Gaussian law of errors. *In other words, the variation of a particular anthropometric characteristic (in this case, height) in a population of individuals is distributed in precisely the same way as the measurement errors that Gauss analyzed, made by astronomers.*

The mean height of a population of soldiers represented something real; the 'true height' of an army recruit. On this account, deviation in height between individuals could simply be treated as noise obscuring an ideal value.

The *average score* is the *true value* of the group under consideration while the deviation from the mean is the result of *accidental causes* that are fundamentally unanalyzable.

A population could either refer to a very broad group of individuals or a very specific group; for example the population of women undergraduates in their second year at this university would be a construct that had as much validity for Quetelet as the construct of an average truck driver in France. The mean of the group with respect to particular measure represented the *idealized type* while the variation was treated simply as a veil that had to be seen through to arrive at this average person. It follows that there could be no science of individual differences; variation was considered to be the result of noise that obscured an ideal type.

Accidental Causes.	random influences that are filtered out by averaging many independent observations
Constant Causes.	always act in same way in a continuous fashion
Variable Causes.	act in a continuous manner, but they vary over time. For example, a cause can change depending on whether it is day or night, or summer versus winter

The greater the number of individuals observed, the more do individual peculiarities, whether physical or moral, become effaced ('effaced' means erased or cancelled out*), and leave in prominent point of view the general facts, by virtue of which society exists and its importance is preserved.*

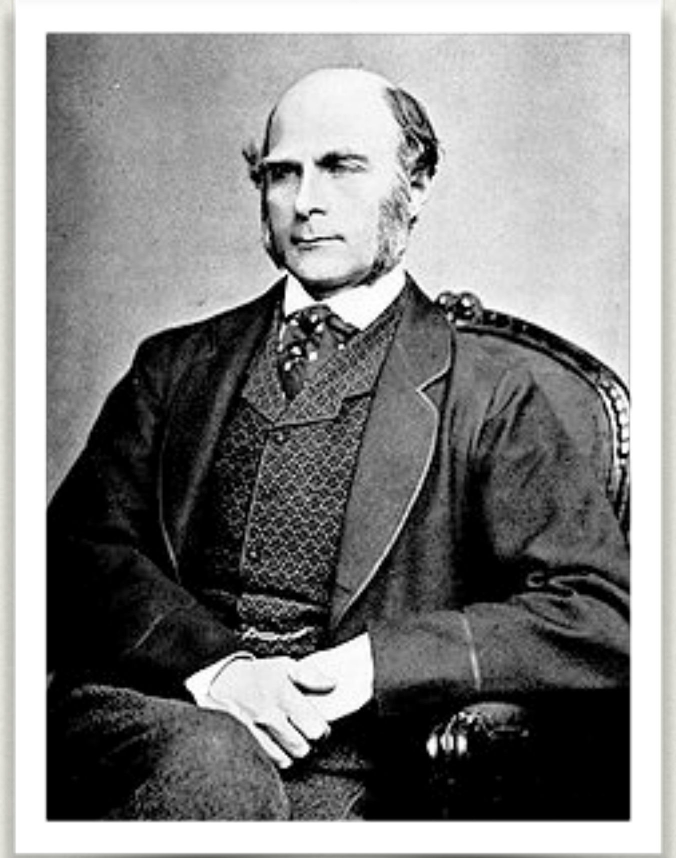
It is the social body, which forms the object of our researchers, and not the peculiarities distinguishing the individuals composing it.

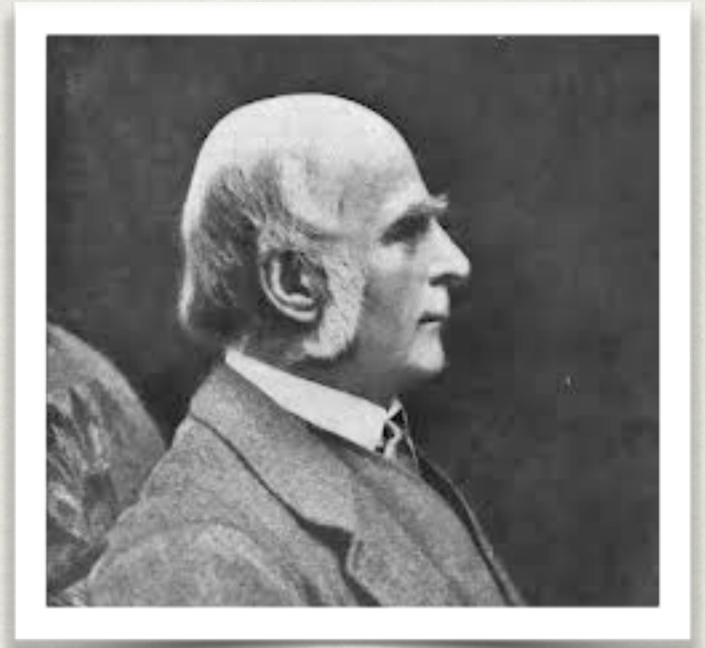
The concept of an average human-being permitted Quetelet to do away with the need to consider *particular* individuals. As he put it: '*It is the social body, which forms the object of our researchers, and not the peculiarities distinguishing the individuals composing it*'. Having drawn this inference, it was very difficult for Quetelet and other social statisticians, to reflect on the importance of the individual or of deviations from an average.

Venn

When we perform an operation ourselves with a clear consciousness of what we are aiming at, we may quite correctly speak about every deviation from this as being an error; but when Nature presents us with a group of objects of every kind, it is a rather bold metaphor to speak in this case also of a law of error, as if she had been aiming at something all the time, and had like the rest of us missed her mark more or less in every instance

GALTON'S BREAKTHROUGH





This was the challenge faced by Galton. As he described the problem in later life, '....*the primary objects of the Gaussian Law of Errors were exactly opposed, in one sense, to those to which I applied them. They were to be got rid of, or to be proved a just allowance for errors. But these errors or deviations were the very things I wanted to preserve and to know about*'.

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.

TABLE I.

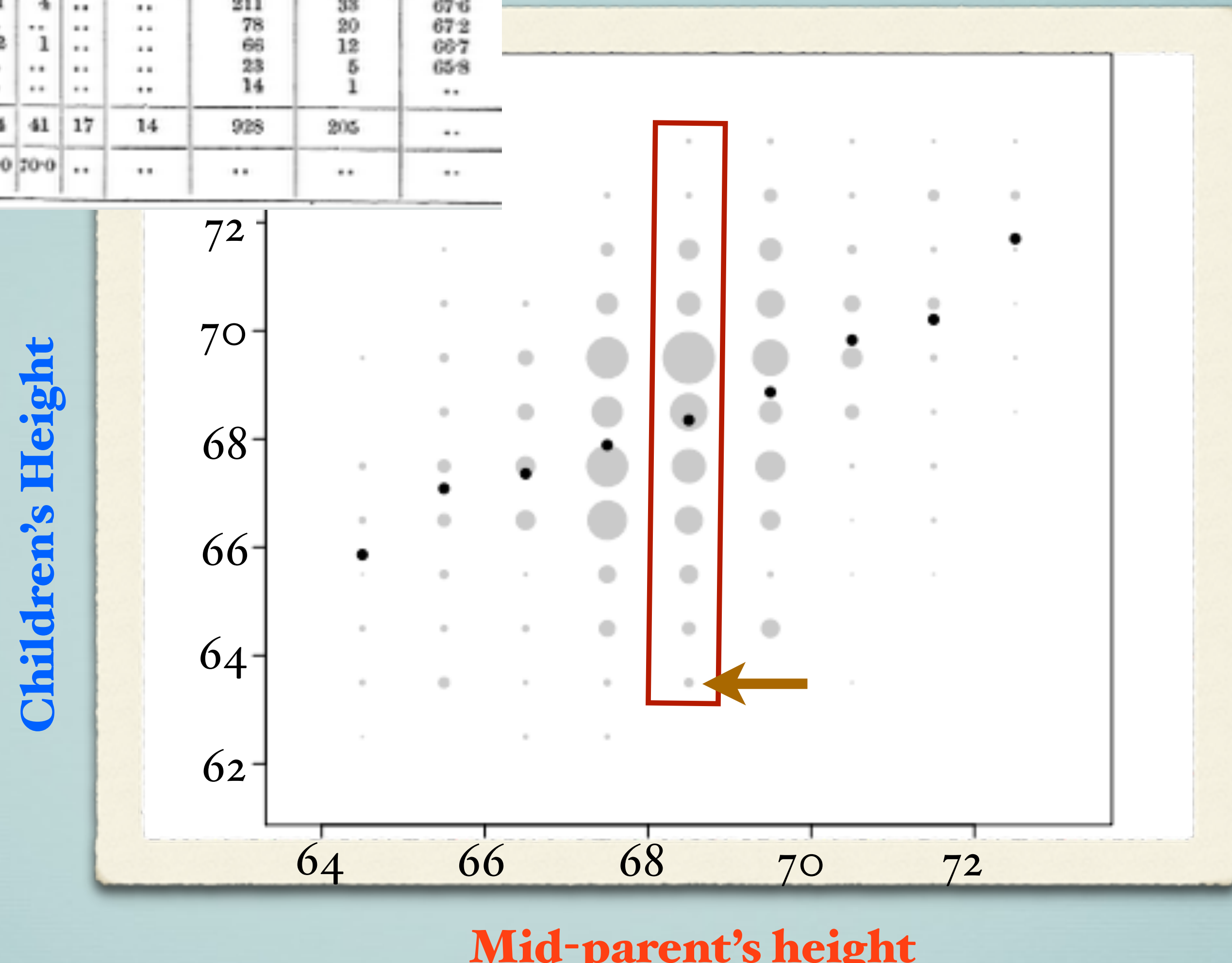
Number of Adult Children of various statures born of 205 Mid-parents of various statures.
(All Female heights have been multiplied by 1·08).

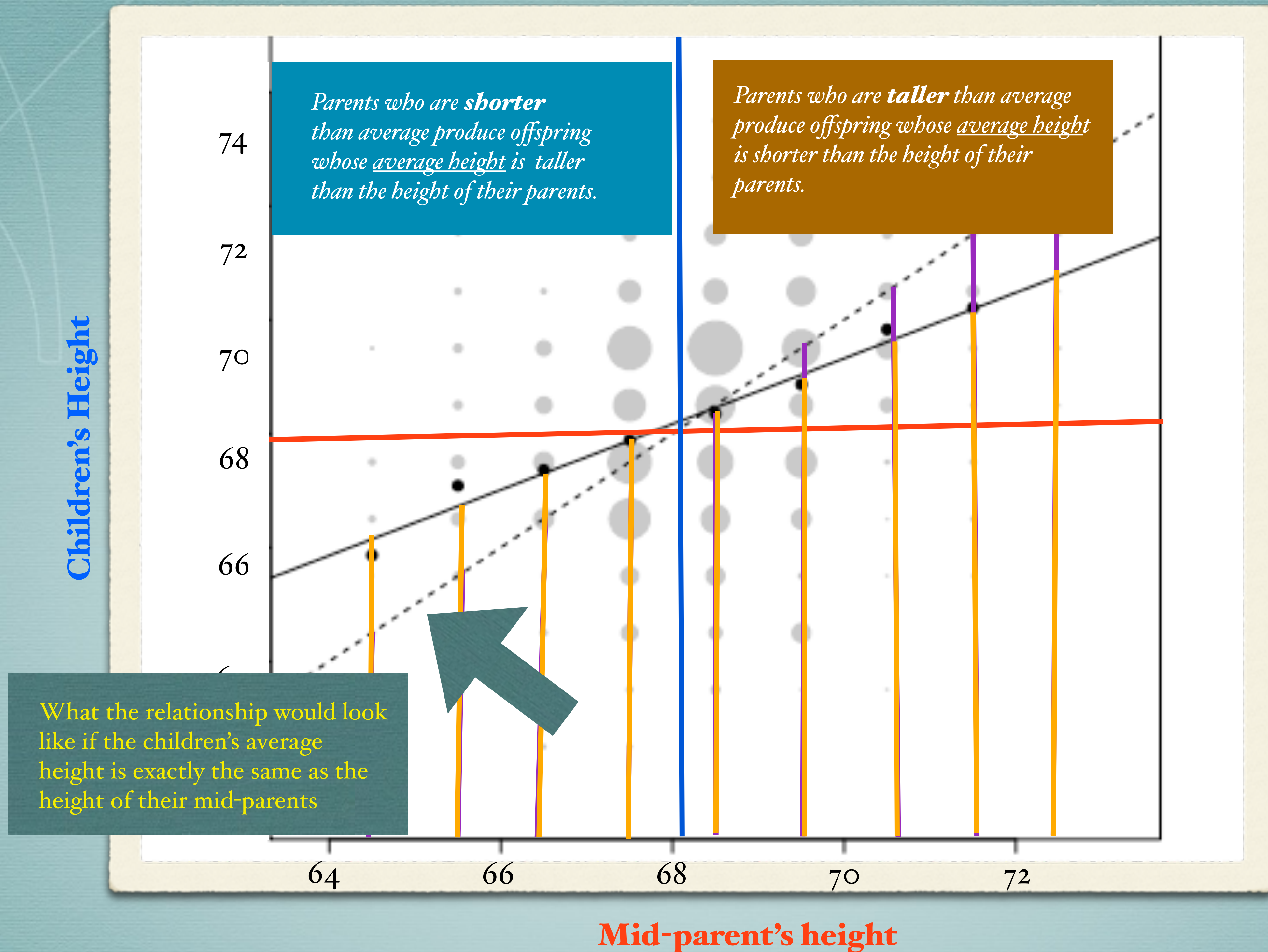
Heights of the Mid-parents in inches.	Heights of the Adult Children.														Total Number of		Medians.
	Below	62·2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above	Adult Children.	Mid-parents.	
Above ..	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	..
72·5	..	..	..	..	..	..	..	1	2	1	2	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	..	..	..	..	..

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	..	..	..	..	..





Children's Height
74
72
70
68
66
Parents who are **shorter** than average produce offspring whose average height is taller than the height of their parents.
Parents who are **taller** than average produce offspring whose average height is shorter than the height of their parents.
What the relationship would look like if the children's average height is exactly the same as the height of their mid-parents
64
66
68
70
72
Mid-parent's height

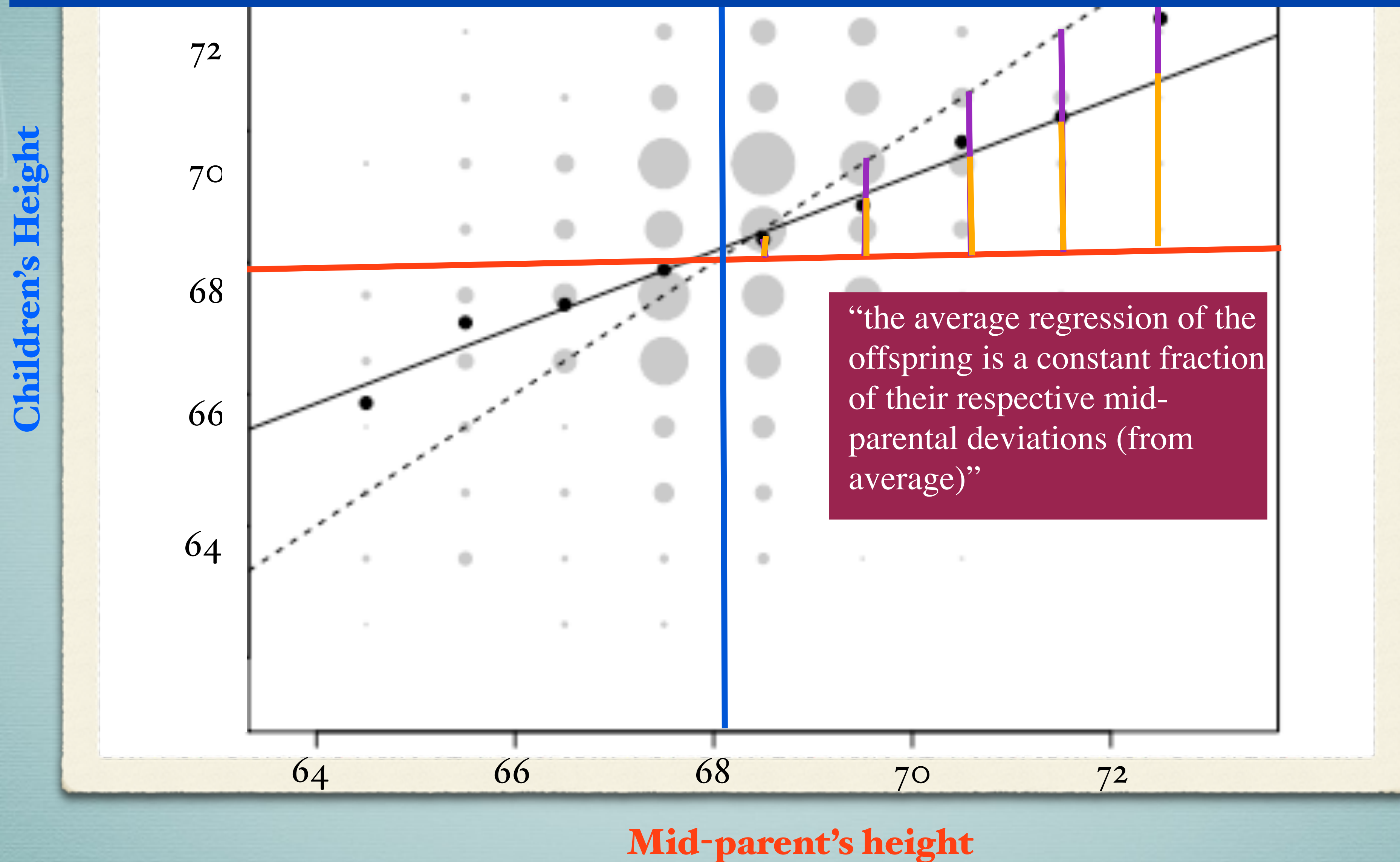
*Parents who are **taller** than average produce offspring whose average height is shorter than the height of their parents.*

*Parents who are **shorter** than average produce offspring whose average height is taller than the height of their parents.*

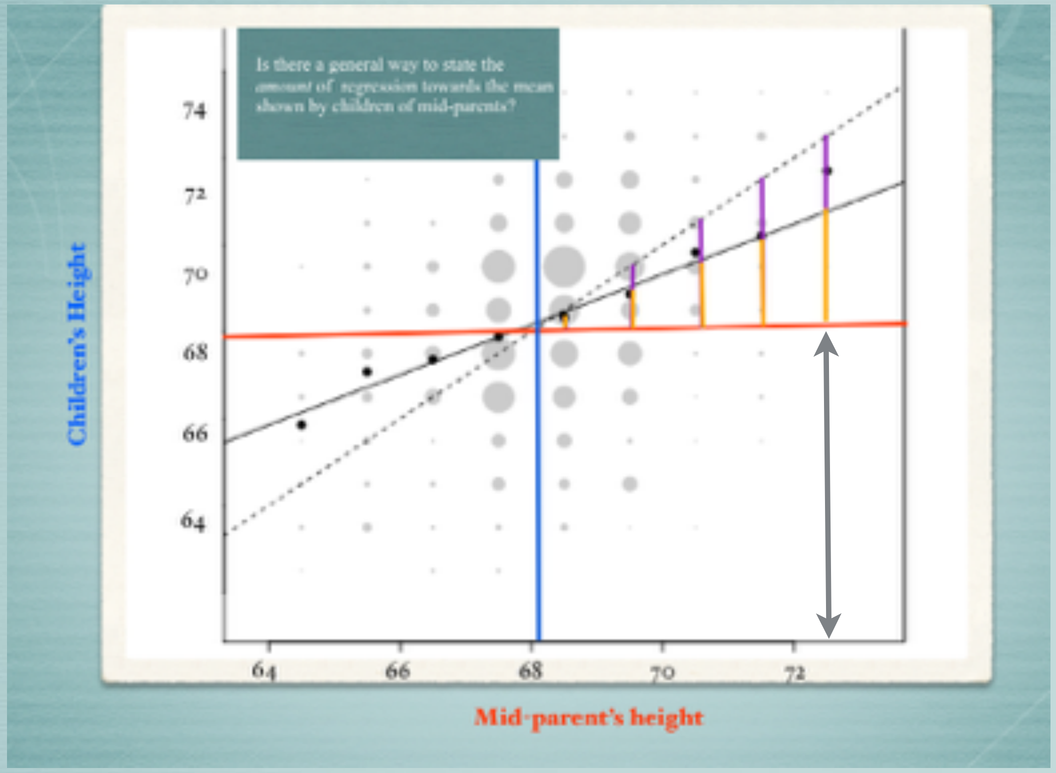
Reversion towards mediocrity

Regression towards the mean

This means that the difference between a child and its parents is proportional to the parents' deviation from average people in the population. If its parents are each *two inches taller than average*, the child will be shorter than its parents by some factor times two inches. For height, Galton estimated this coefficient to be about 2/3: the height of an individual will -- on the average --be two thirds of the parents’ deviation from the population average.

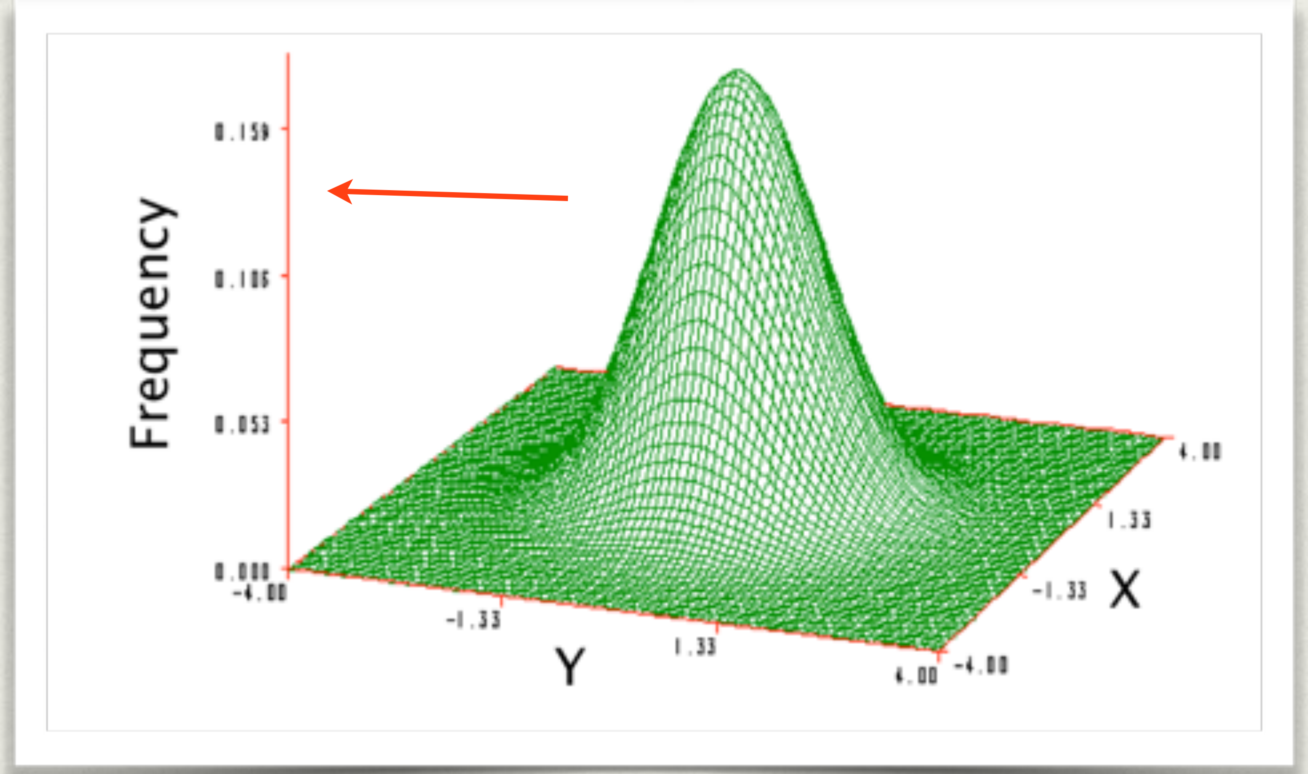


“the average regression of the offspring is a constant fraction of their respective mid-parental deviations... from average”



“the average regression of the offspring is a constant fraction of their respective mid-parental height”

A bivariate normal distribution

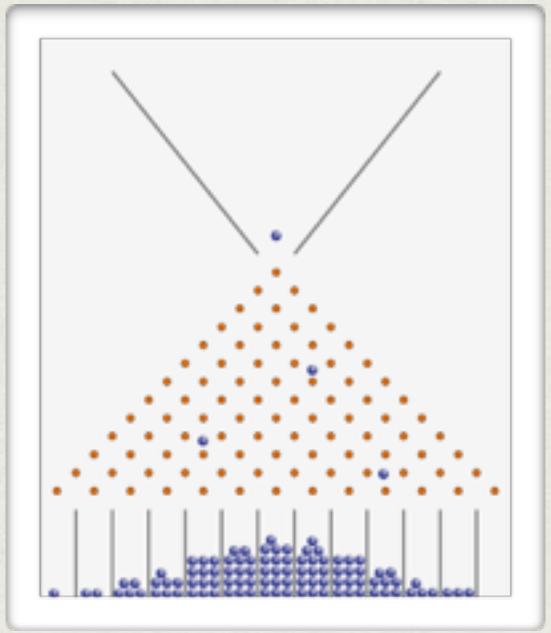


Galton’s explanation of regression towards the mean in children’s height.

A child inherits partly from his parents, partly from his ancestors. Speaking generally, the further his genealogy goes back, the more numerous and varied will his ancestry become, until they cease to differ from any equally numerous sample taken at haphazard from the race at large.

The modern explanation

Father	Mother	Child
+ +	+ +	+ +
- -	+ +	- +
+ +	- -	+ -
+ +	+ -	+ -
+ +	+ +	+ +
- -	+ +	- +
+ +	+ +	+ +
5	5	3



You can see a modern copy of Galton's Quincunx in action by clicking on the word in this sentence.

Parents

Children

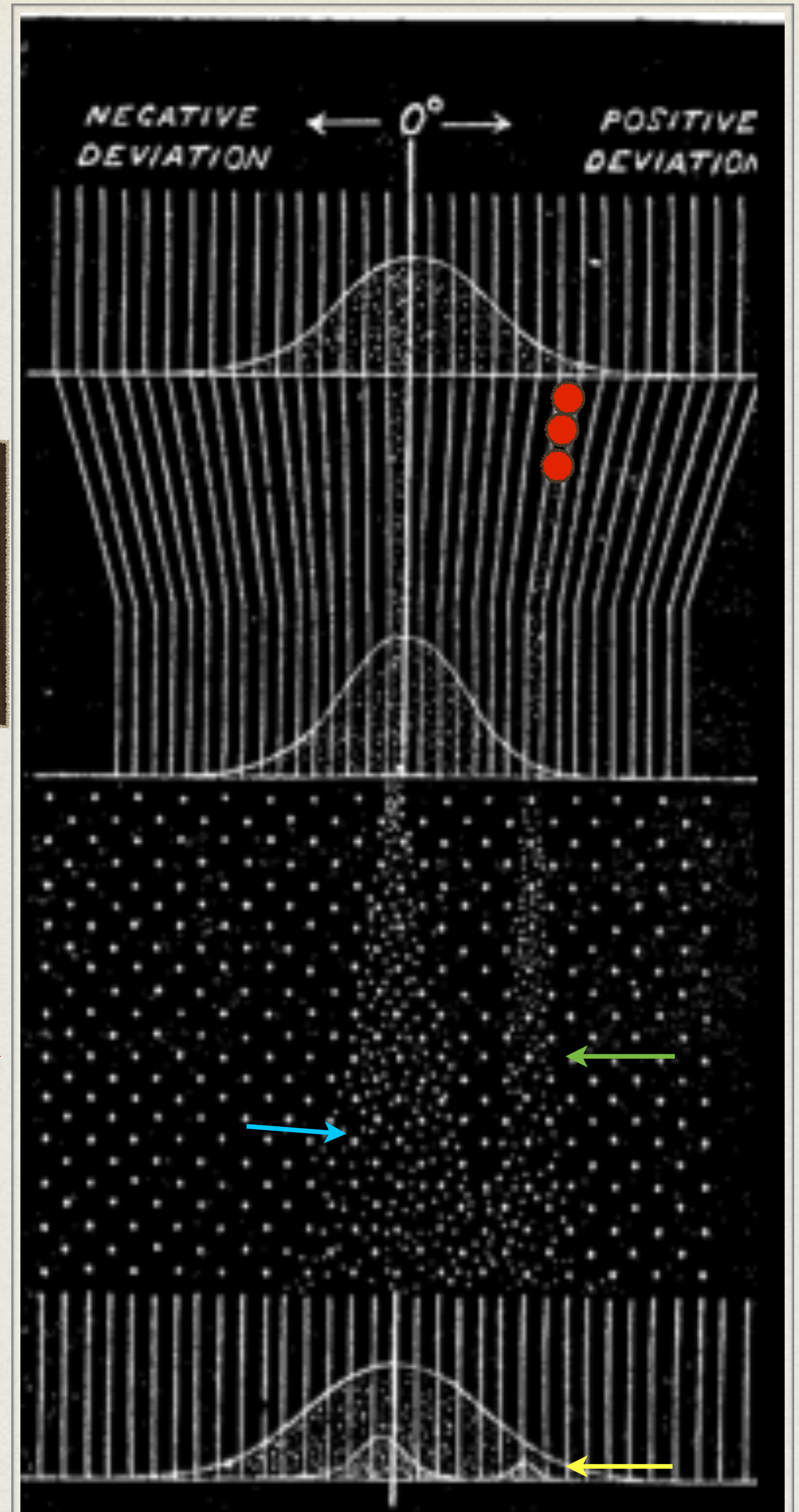


TABLE I.

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	..	..	..	..	..