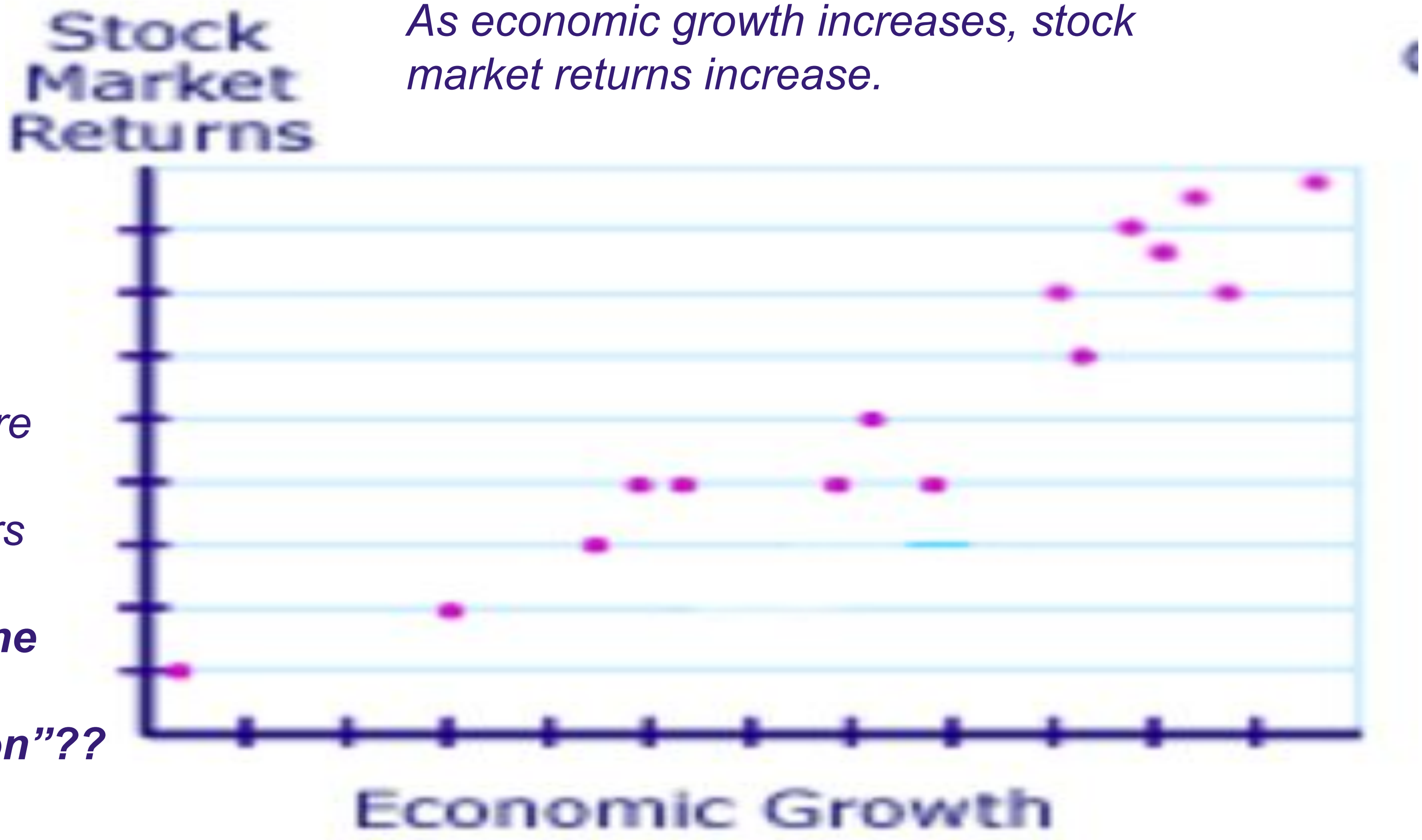
Variance, Co-Variance, Mahalanobis distance, Co-Relation, Mean, Standard Deviation

Prepared By: Dr.Mydhili K Nair, Professor, ISE Dept, RIT
For: Machine Learning Elective Class
Target Audience: Sem 6 Students
Term: Feb to June 2019

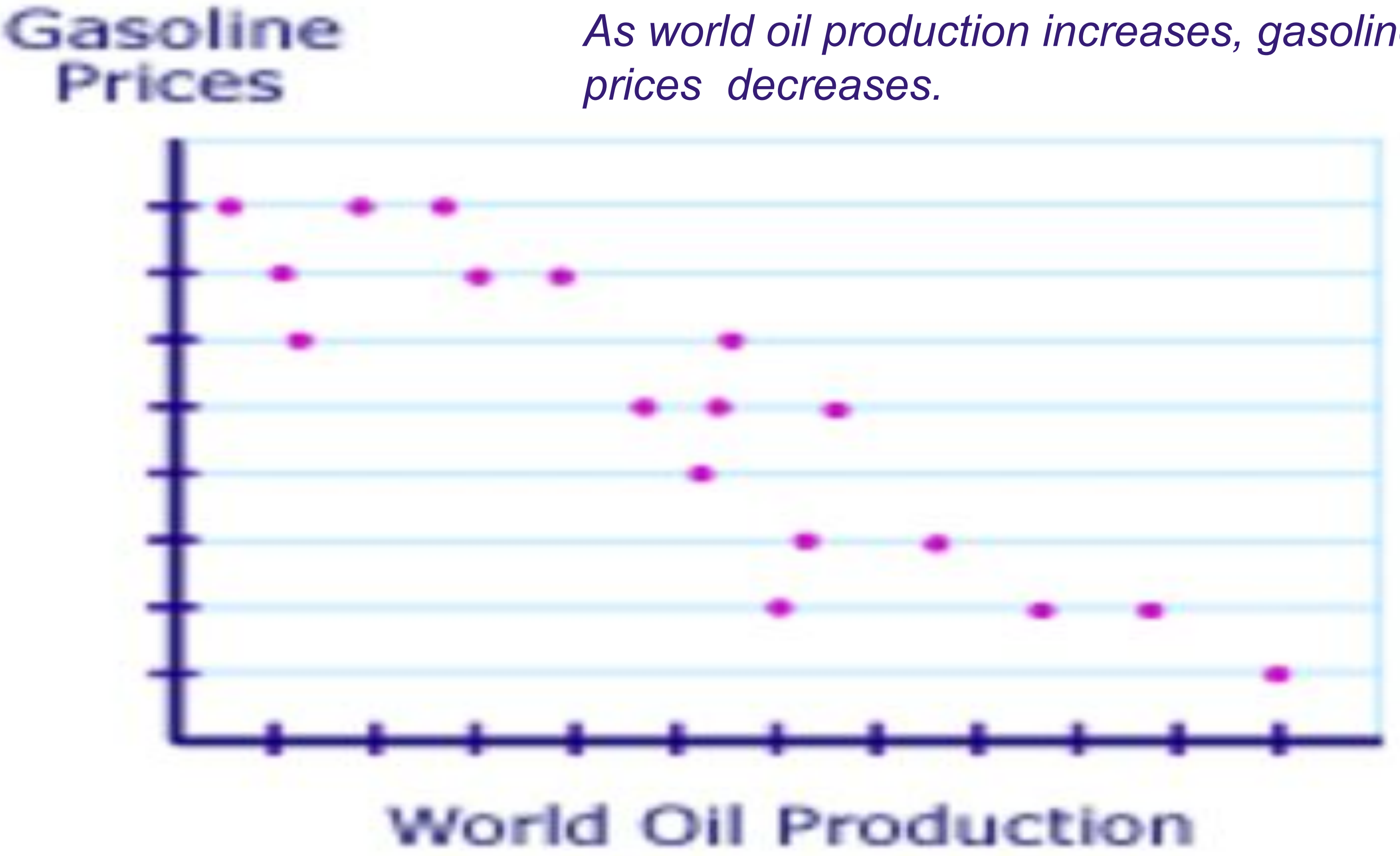
As economic growth increases, stock market returns increase.



By what quantity are these two parameters related?

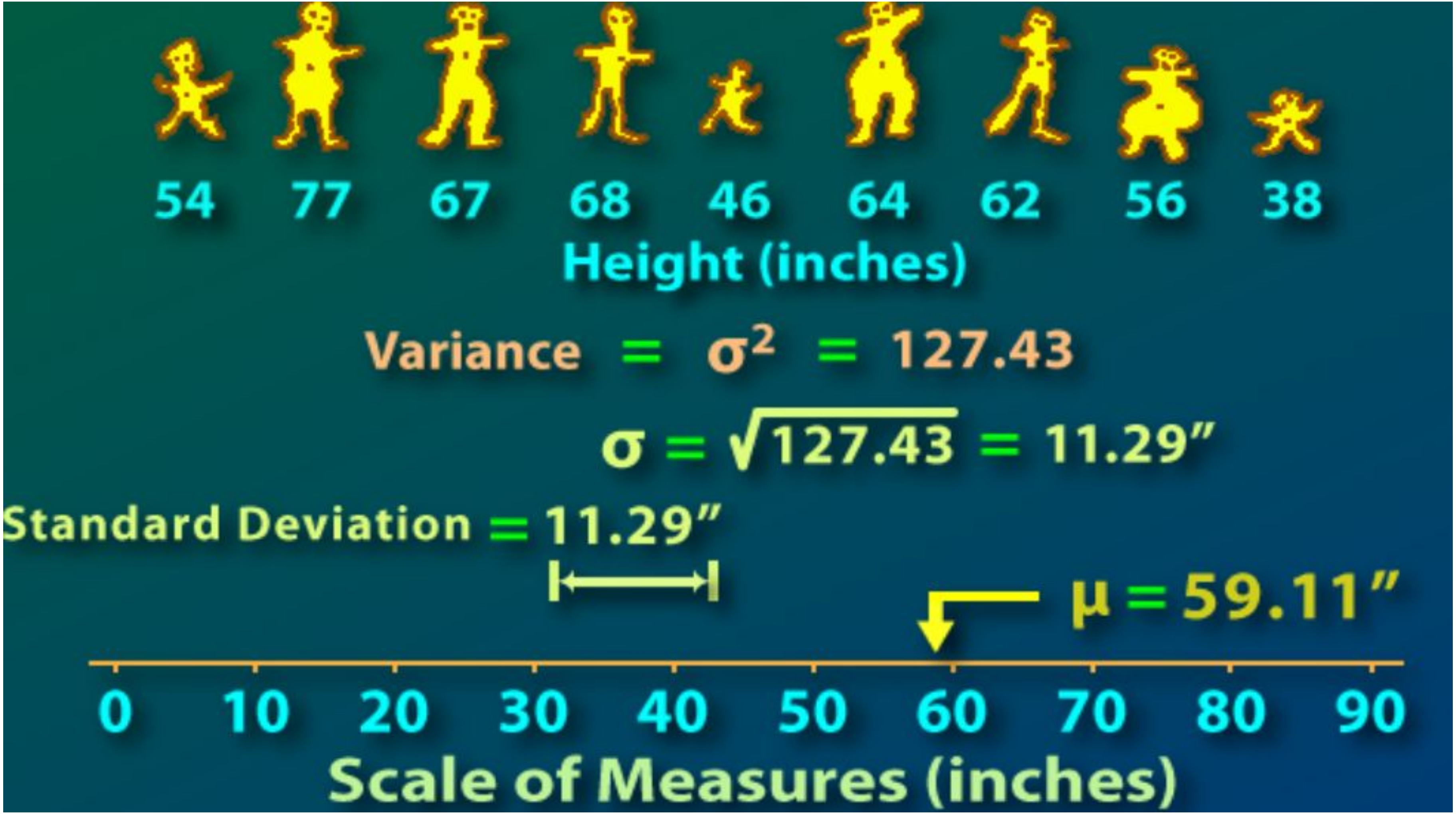
What is the "level of correlation"??

As world oil production increases, gasoline prices decreases.



By what quantity are these two parameters related?

***What is the "level of correlation"**??*


54 77 67 68 46 64 62 56 38
Height (inches)
Variance = σ² = 127.43
σ = √127.43 = 11.29″
Standard Deviation = 11.29″
μ = 59.11″
0 10 20 30 40 50 60 70 80 90
Scale of Measures (inches)

Formulae

$$Mean = \bar{x} = \frac{\sum x}{n}$$

$$Variance = s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

$$Std\ dev = s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

<u>Note</u>: The numerator is **not** a squared deviation, but a **product** of the summation deviation of each data point X from its mean and each data point Y from its mean.

Covariance

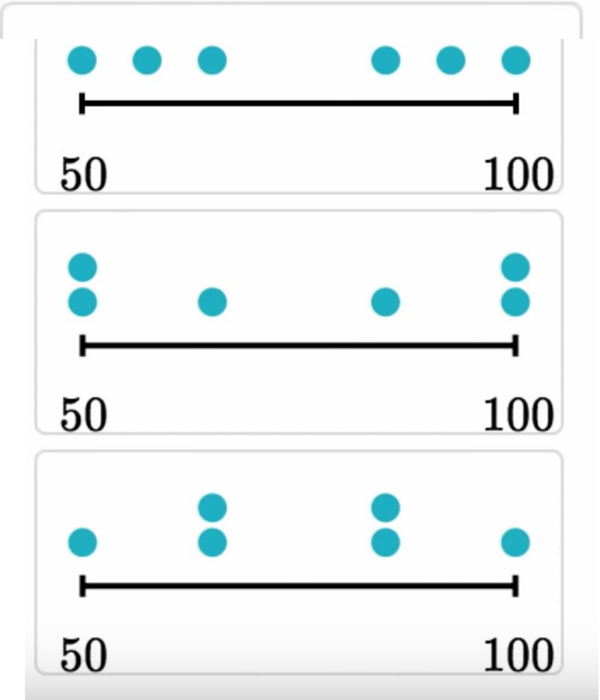
$$COV(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Correlation Coefficient

$$r_{(x,y)} = \frac{COV(x,y)}{Std\ dev}$$

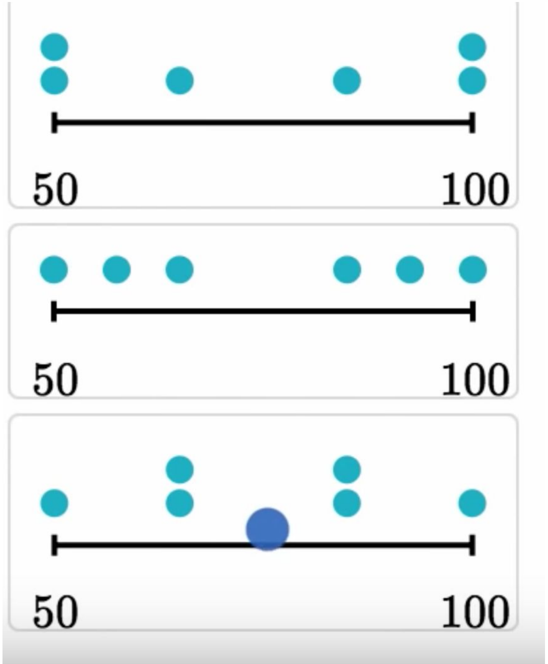
Each dot plot below represents a different set of data.

Order the dot plots from largest standard deviation (top) to smallest standard deviation (bottom).



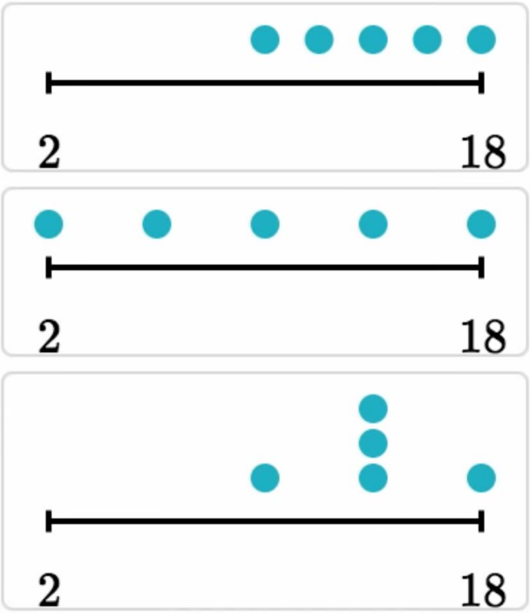
Standard Deviation in Visual Form
E.g.#1

Answer



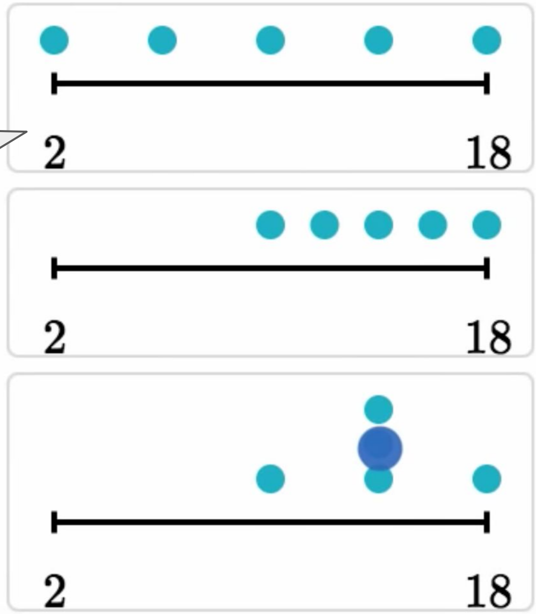
Each dot plot below represents a different set of data.

Order the dot plots from largest standard deviation (top) to smallest standard deviation (bottom).



Standard Deviation in Visual Form
E.g.#2

Answer



σ_X

Question#1 : Standard Deviation Calculation

A veterinarian weighed a sample of 6 puppies. Here are each of their weights (in kilograms):

$$1, 2, 7, 7, 10, 15$$

The mean of these weights is $\bar{x} = 7 \text{ kg}$.

What is the standard deviation?

Standard Deviation in Visual Form E.g.#3

Question#1 : Standard Deviation Calculation

x_i	Deviation: $(x_i - \bar{x})$	Squared deviation: $(x_i - \bar{x})^2$
1	$1 - 7 = -6$	$(-6)^2 = 36$
2	$2 - 7 = -5$	$(-5)^2 = 25$
7	$7 - 7 = 0$	$0^2 = 0$
7	$7 - 7 = 0$	$0^2 = 0$
10	$10 - 7 = 3$	$3^2 = 9$
15	$15 - 7 = 8$	$8^2 = 64$
Sum:	0	134

Standard Deviation in Visual Form E.g.#3

Question#1 : Standard Deviation Calculation

$$\frac{134}{n-1} = \frac{134}{6-1} = \frac{134}{5} = 26.8$$

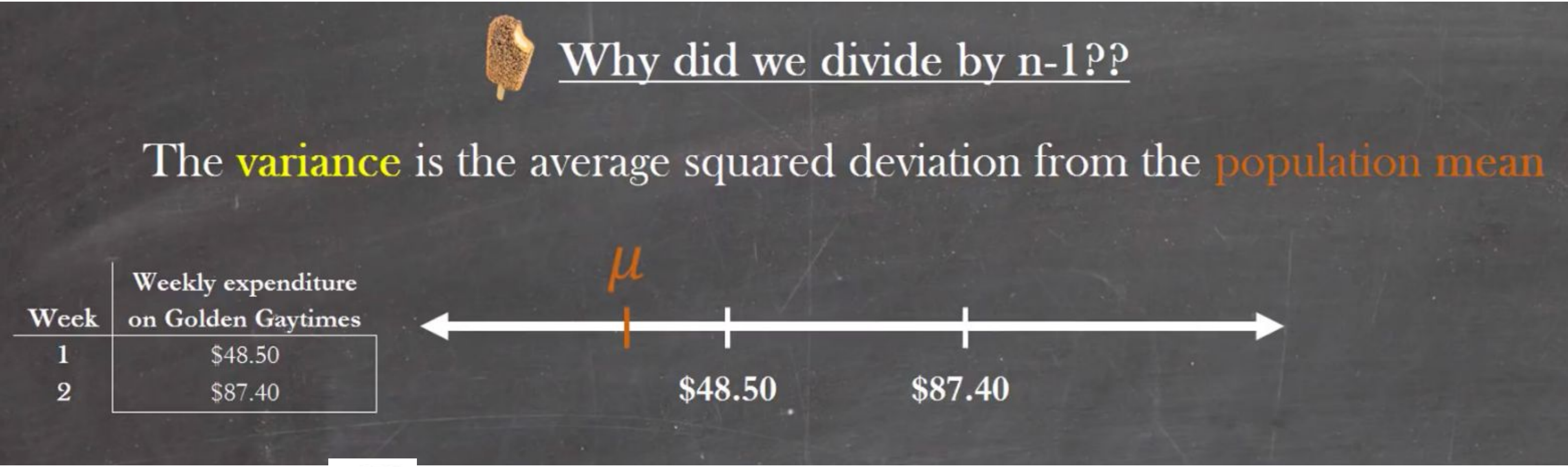
$$s_x = \sqrt{26.8} \approx 5.177 \text{ kg}$$

$$Variance = s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

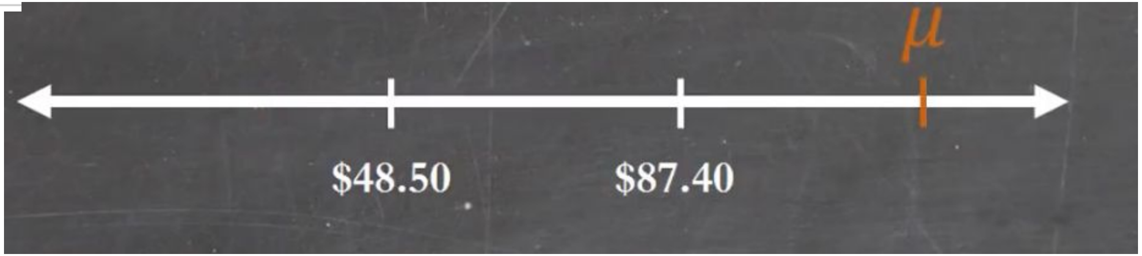
$$Std\ dev = s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

Standard Deviation in Visual Form
E.g.#3

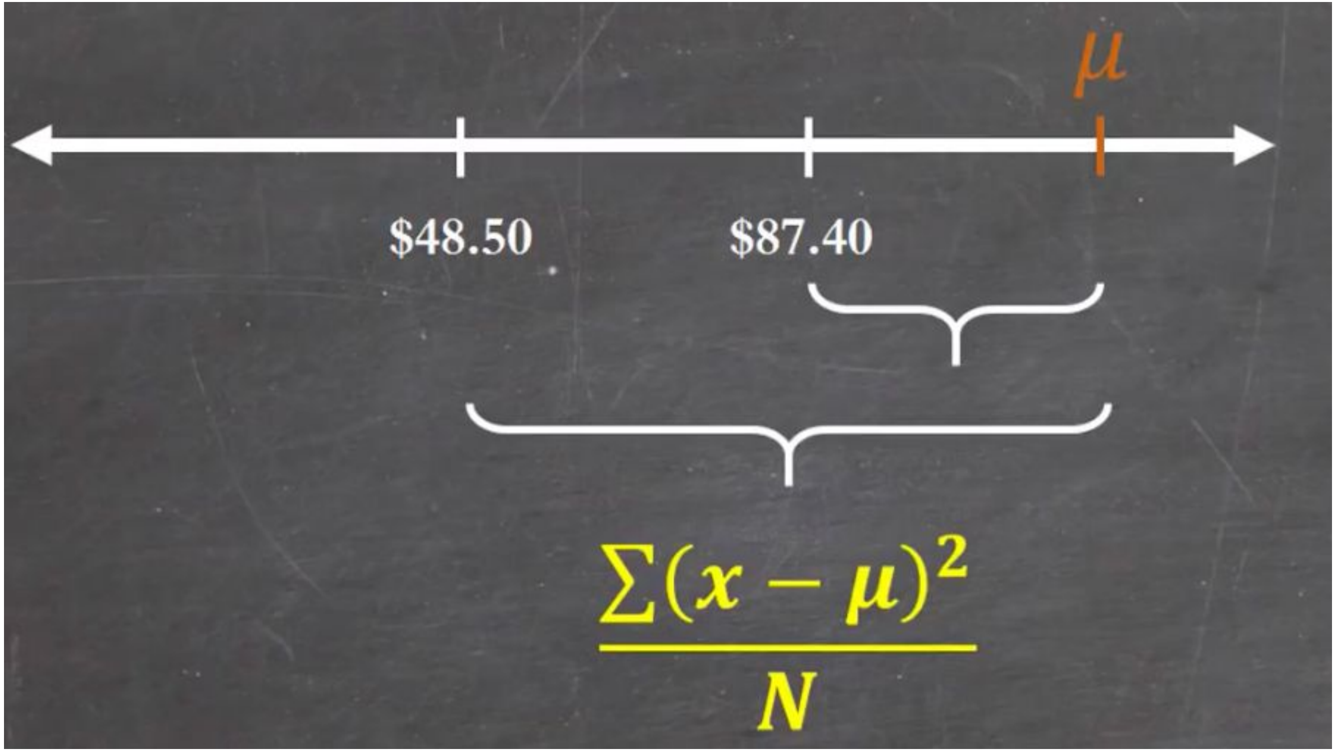
Standard Deviation Formula Explanation
(Why did we divide by n-1 ?)



Population Mean μ could be anywhere on this number line.

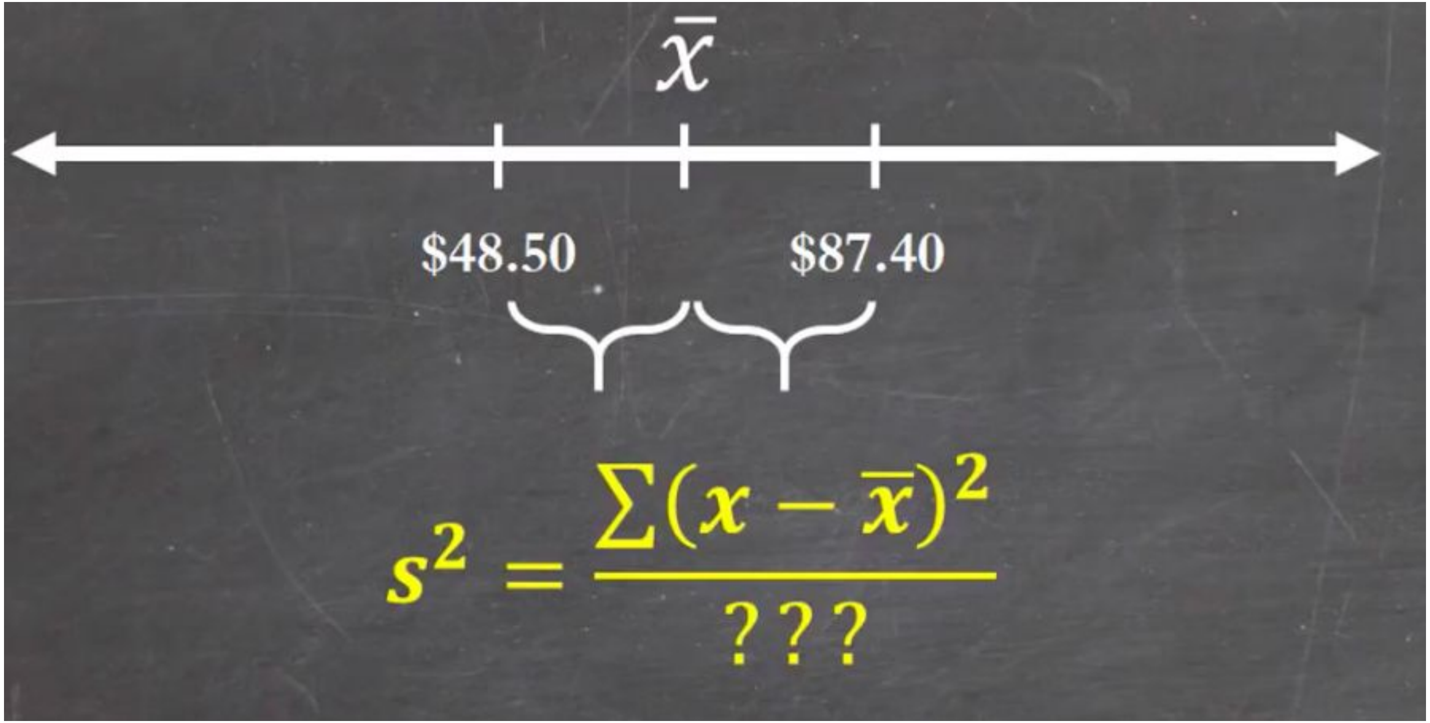


Standard Deviation Formula Explanation
(Why did we divide by n-1 ?)



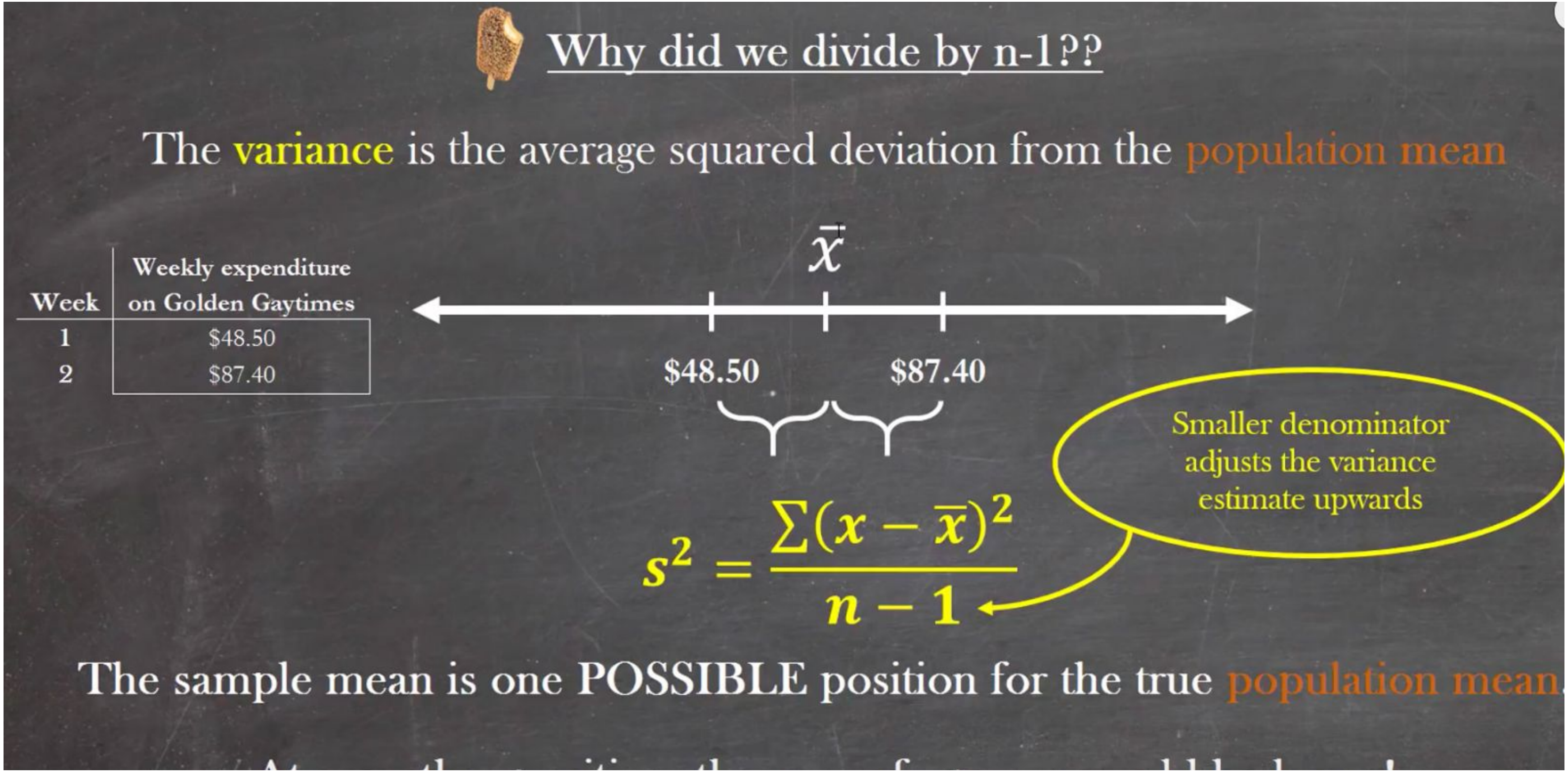
This is the **Population Variance**

Standard Deviation Formula Explanation
(Why did we divide by n-1 ?)



This is the **Sample Variance**

Standard Deviation Formula Explanation
(Why did we divide by n-1 ?)

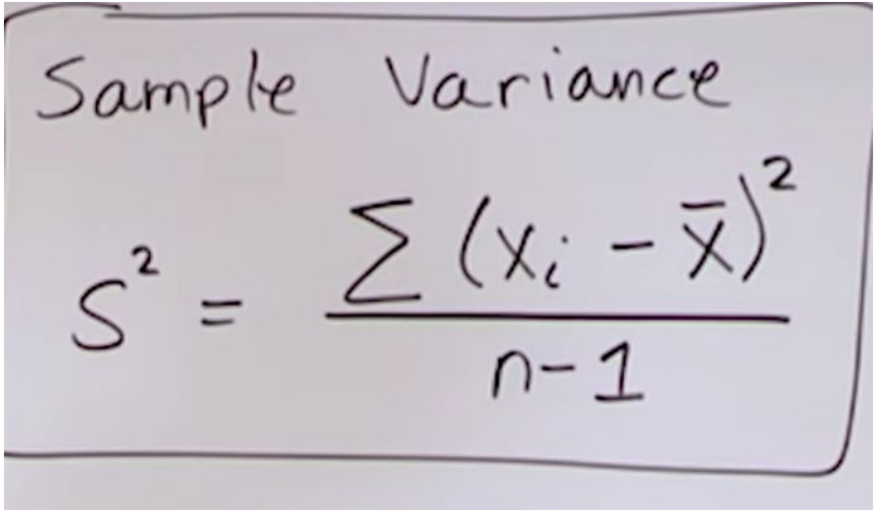


σ^2

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Population Variance

Difference between Population and Sample Variance



How "connected" or "correlated" are these two parameters?

Sl#	Temperature	Ice-Cream Sales
1	66	8
2	72	11
3	77	15
4	84	20
5	83	21
6	71	11
7	65	8
8	70	10
Mean	73.5	13
Std Deviation	7.19	5.13

How “connected” or “correlated” are these two parameters?

Sl#	Deviation $(x - \bar{x})$	Sample Variance $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
1	-7.5	56.25
2	-1.5	3
3	3.5	12.25
4	10.5	110.25
5	9.5	90.25
6	-2.5	5
7	-8.5	17
8	-3.5	12.25

$(x - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
-7.5	-5	37.5
-1.5	-2	3
3.5	2	7
10.5	7	73.5
9.5	8	76
-2.5	-2	5
-8.5	-5	42.5
-3.5	-3	10.5
		255

$$COV(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

= 255 / 7

= **36.43**

Covariance is a large positive number. This indicates *positive correlation.*

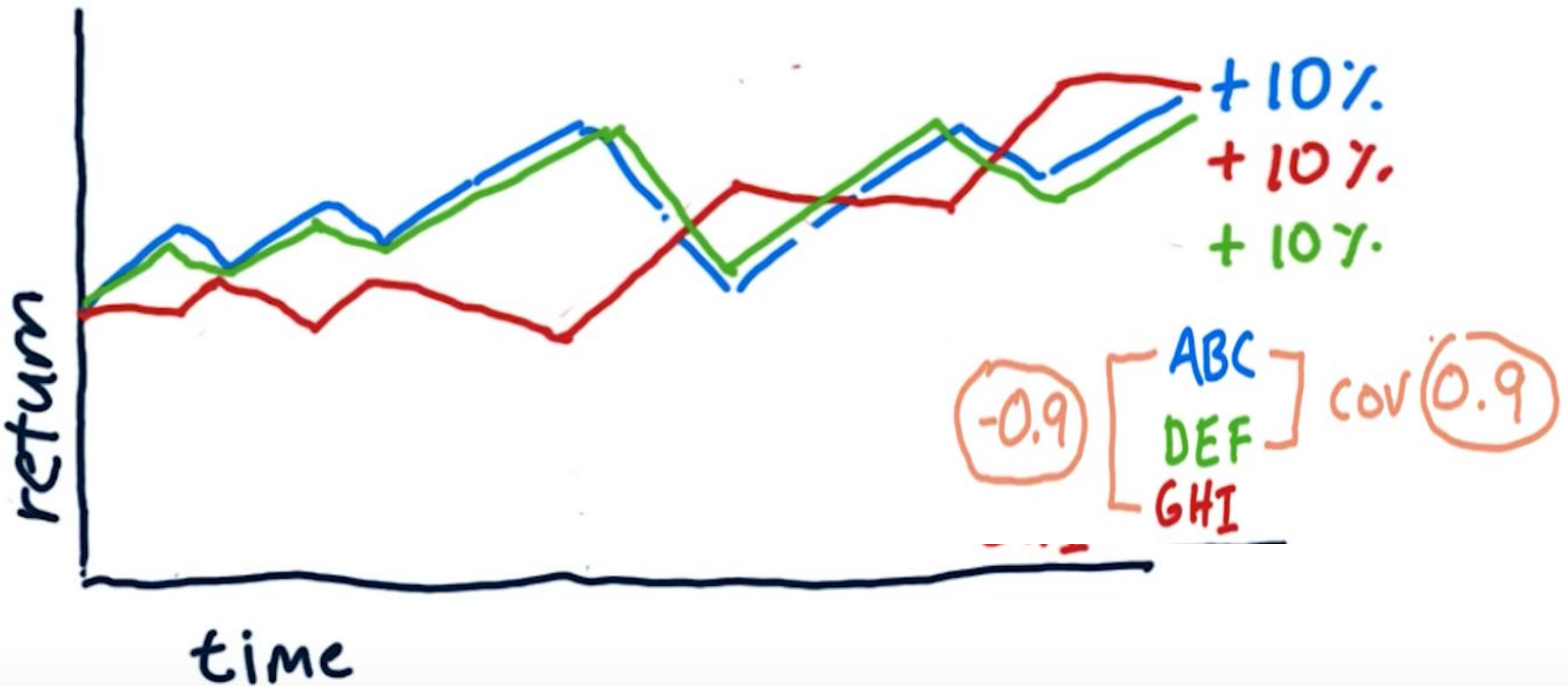
$$r_{(x,y)} = \frac{COV(x,y)}{s_x s_y}$$

= 255 / (7.19 * 5.13)

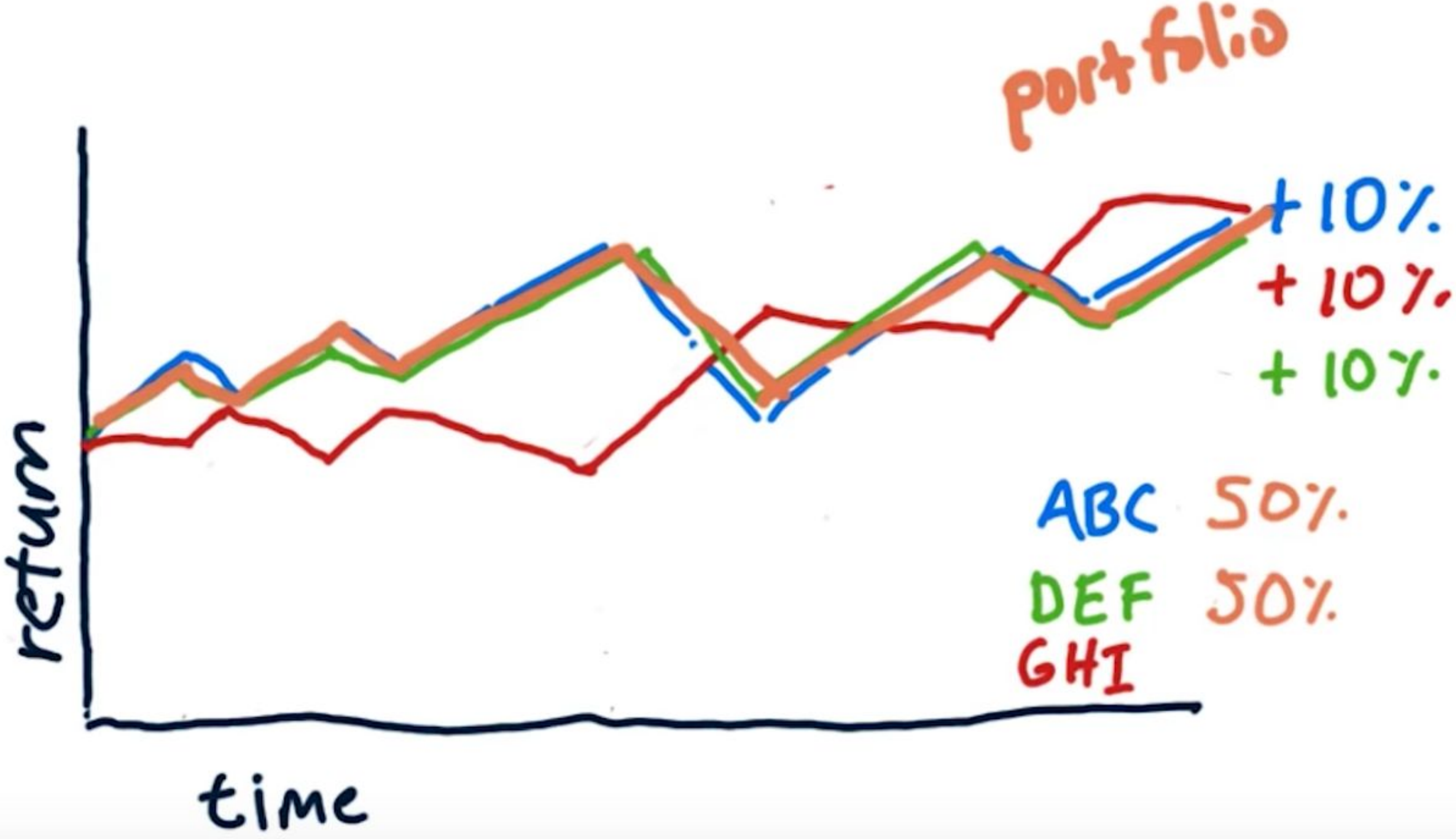
= **0.99**

Correlation Coefficient is number close to 1. This indicates *correlation is very high between X and Y. .*

The importance of covariance



The importance of covariance



The importance of covariance