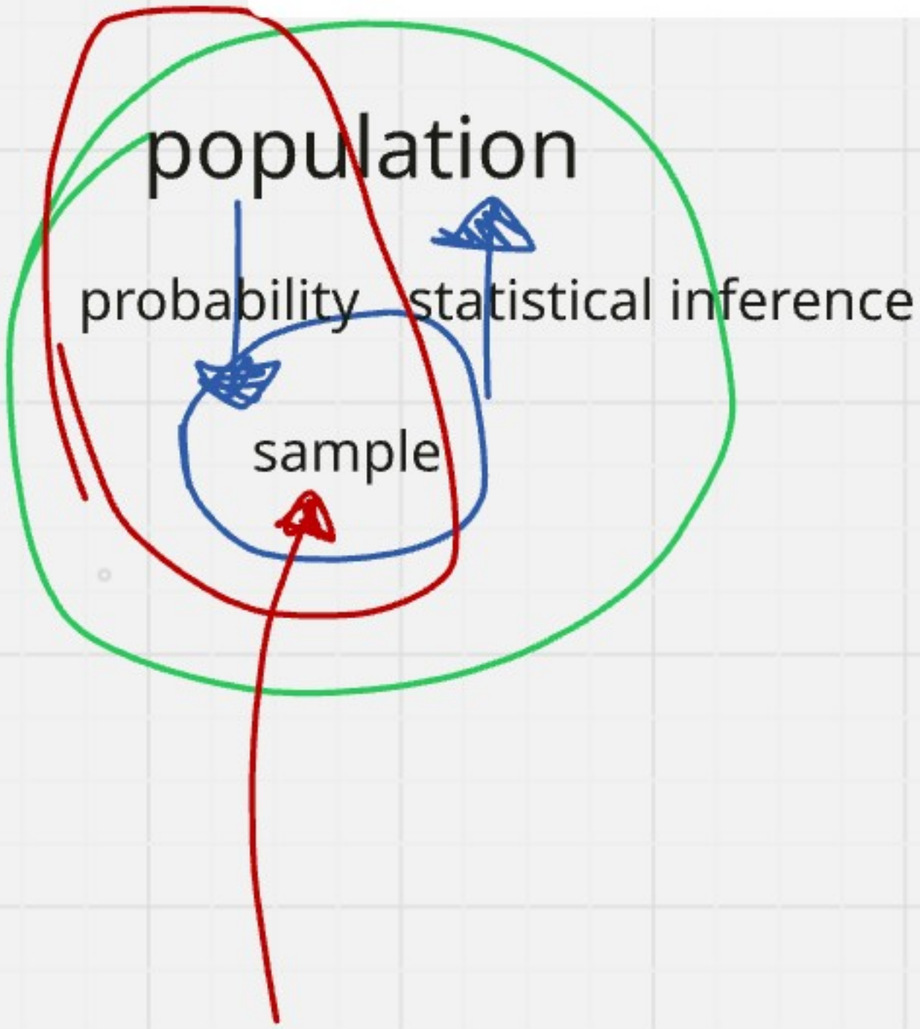


Variance! (varies means changes); in particular, variance is the squared distance (deviation) between each variable to the mean

- Dependent Variable - y
- Mean = Average
- intuition: a very high variance means the y changes a lot!!
    - what would a very low variance mean? Ans: It doesn't change a lot!!

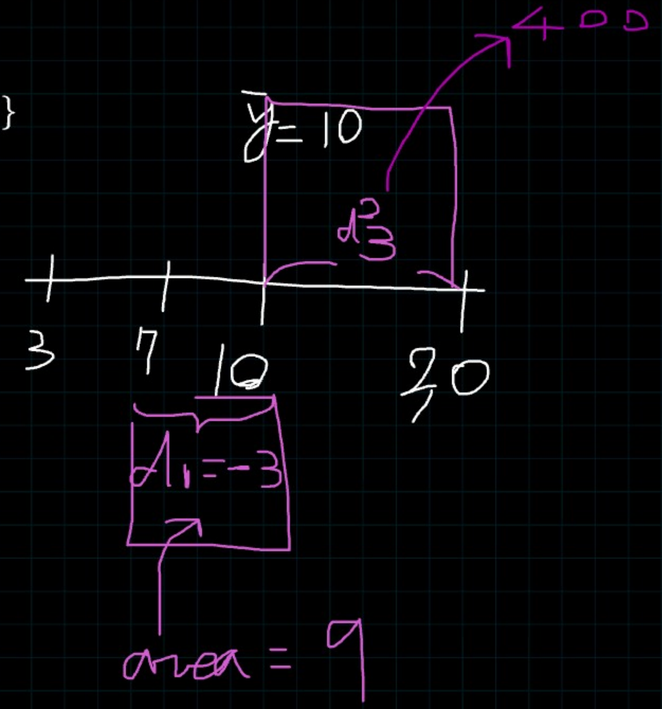
$\sum_{i=1}^{n} a_i := a_1 + \cdots + a_n$
$\sum_{i=1}^{4} a_i := a_1 + a_2 + a_3 + a_4$

We call $\sum$ a sum! or we call it 'sigma'
$\bar{y}$ = **mean**
y_i = dependent variable

$\frac{1}{n-1}(\sum_{i=1}^{n}(y_i - \bar{y})^2) = \text{Variance of } y = Var(y) = S^2$

Say we have a set of data consisting of $y_i, \{7, 10, 20, 3\}$
In this set of data, $\bar{y} = 10$
and for $y_i, y_1 = 7, y_2 = 10, y_3 = 20, y_4 = 3$
To find the variance we first find the distance!
By distance we mean the 'deviation':

$(y_1 - \bar{y})^2 = (-3)^2 = 9 := deviation_1^2$
$(y_2 - \bar{y})^2 = (0)^2 = 0 := deviation_2^2$
$(y_3 - \bar{y})^2 = (10)^2 = 100 := deviation_3^2$
$(y_4 - \bar{y})^2 = (-7)^2 = 49 := deviation_4^2$



When we add everything togeter, we get
$\sum_{i=1}^{4}(y_i - \bar{y})^2 = 9 + 0 + 100 + 49 = 158$
Since n = 4, $\frac{1}{n-1} = \frac{1}{3}$.

So!! Variance of y = $\frac{1}{3}\sum_{i=1}^{4}(y_i - \bar{y})^2 = 158/3 = 52.6$

WHY would we square the whole thing!!??
Lets do an experiment and we shall see!
Say we have the same set of data consisting of $y_i, \{7, 10, 20, 3\}$
Now we know that $\bar{y} = 10$.

$(y_1 - \bar{y}) = deviation_1 = 7 - 10 = -3$
$(y_2 - \bar{y}) = deviation_2 = 10 - 10 = 0$
$(y_3 - \bar{y}) = deviation_3 = 20 - 10 = 10$
$(y_4 - \bar{y}) = deviation_4 = 3 - 10 = -7$

Consider adding them together, we get $-3 + 0 + 10 + (-7) = 0$.

This is exactly why we square the deviation!!! Because for any data set,
the sum of all deviation always equal to zero; thus it tells us NOTHING!
IT does not provide us with information about the magnitude of changes on the variables!

Expected value E(x) = $\sum_{i=1}^{n} x_i \cdot p(x_i)$
$= x_1 \cdot p(x_1) + x_2 \cdot p(x_2) \cdots + x_{n-1} \cdot p(x_{n-1}) + x_n \cdot p(x_n)$
where x_i is outcome, and $p(x_i) :=$ probability of x_i
Example!!
Rolling a dice. Outcome set is { 1, 2, 3, 4, 5, 6} := x_i;
We can infer that, suppose the dice fair, the probability set is {1/6, 1/6, 1/6, 1/6, 1/6, 1/6}
Note that here order matters for the set!!
And so we can conclude that the expected value for rolling a dice is:
$1 \cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 = 7/2 = 3.5.$
We call this 7.5 the expected value

Question: (1) What is the relationship between Variance and Standard Deviation?
(2) -- Challenge question -- Why would we square the 'Deviation,' and consider variance?

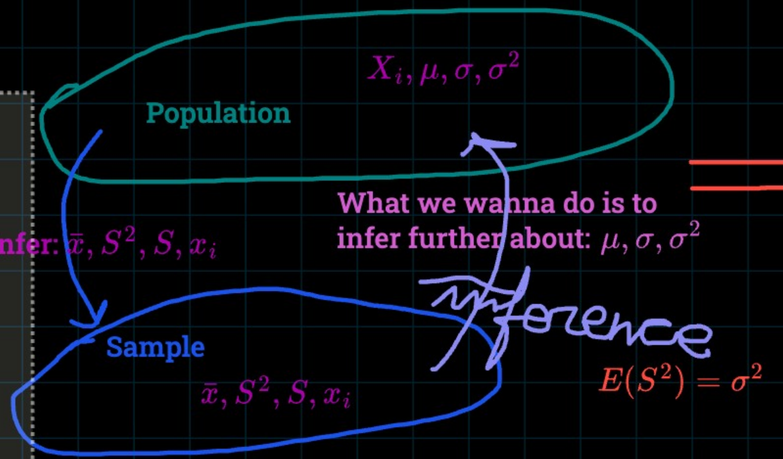
Answer: (1) The Standard Deviation is the square root of the Variance!

Answer: (2) Because no matter what data we choose,
the sum of all the Deviation will always be zero, no exceptions!
That would not help us with information on any changes with the variables in the data.
On the other hand, if we square the Deviation, then the outcome would rarely be zero;
only when all the values in the data are the same. But mostly, the results would be positive numbers and not zero.
Therefore you can get the variance easily without struggling!

Comment: This is absoluetly correct!!! 10/10!!

Yes!!! In particular, for sample

To Answer (1):
This is excellent! Note that
variance $:= S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
we interpret S^2 **as the 'average squared deviation',**
i.e., how it changes.
But!! Note that, we denote variance of a population as
$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$**; where** μ **is the population mean**
NOTE the $\frac{1}{n}$ $vs.$ $\frac{1}{n-1}$ **difference when we**
'average out' population or sample variance.
For the time being we will skip the proof.
But, the reason we define S^2 **as variance** $:= S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
is because MATHEMATICALLY, $E(S^2) = \sigma^2$**,**
That is, the expected value of our sample variance formula gives exactly the population variance which is of our interest of investigation



We can make claims regarding
the population!! Given formula of S.
We call $\frac{1}{n-1}$ the bias accountant!!
It accounts for the bias of sample
to make its expected value equals to the
actual, i.e., σ^2 parameter in the population

Gauss-Markov Assumptions (for Classical Linear Model):

1. Linear in parameters:

$\hat{y} = \beta_0 + \beta_1 x$ **(simple regression model),**

where $\beta_0 = \overline{y} - \beta_1 \overline{x}$**; where 'overline' refers to 'mean';**

and $\beta_1 = \frac{S_{xy}}{S^2 x}$**. Where** $S_{xy} = \text{cov}(xy)$ **is the covariance and the** S^2**x is the variance of** x**.**

2. Random Sampling (important) - I.I.D, i.e., Individually identically distributed random sampling:

we assume that there are n **random observations from the linear model we come up with. Following from this assumption we have**

$\mathbb{E}(\bar{x}) = \mu$**. This means on average out sample mean is the population mean.**

3. No perfect collinarity **If we have a model** $\hat{y} = \beta_0 + \beta_1 x + \beta_2 z$**,**

- no variable is constant **then** x **and** z **are not exactly linearly correlated and not constant**

-no exact linear relationship between two independent variables.

4. Zero Conditional Mean:

$\mathbb{E}(\mu | x_1, \ldots, x_n) = 0$

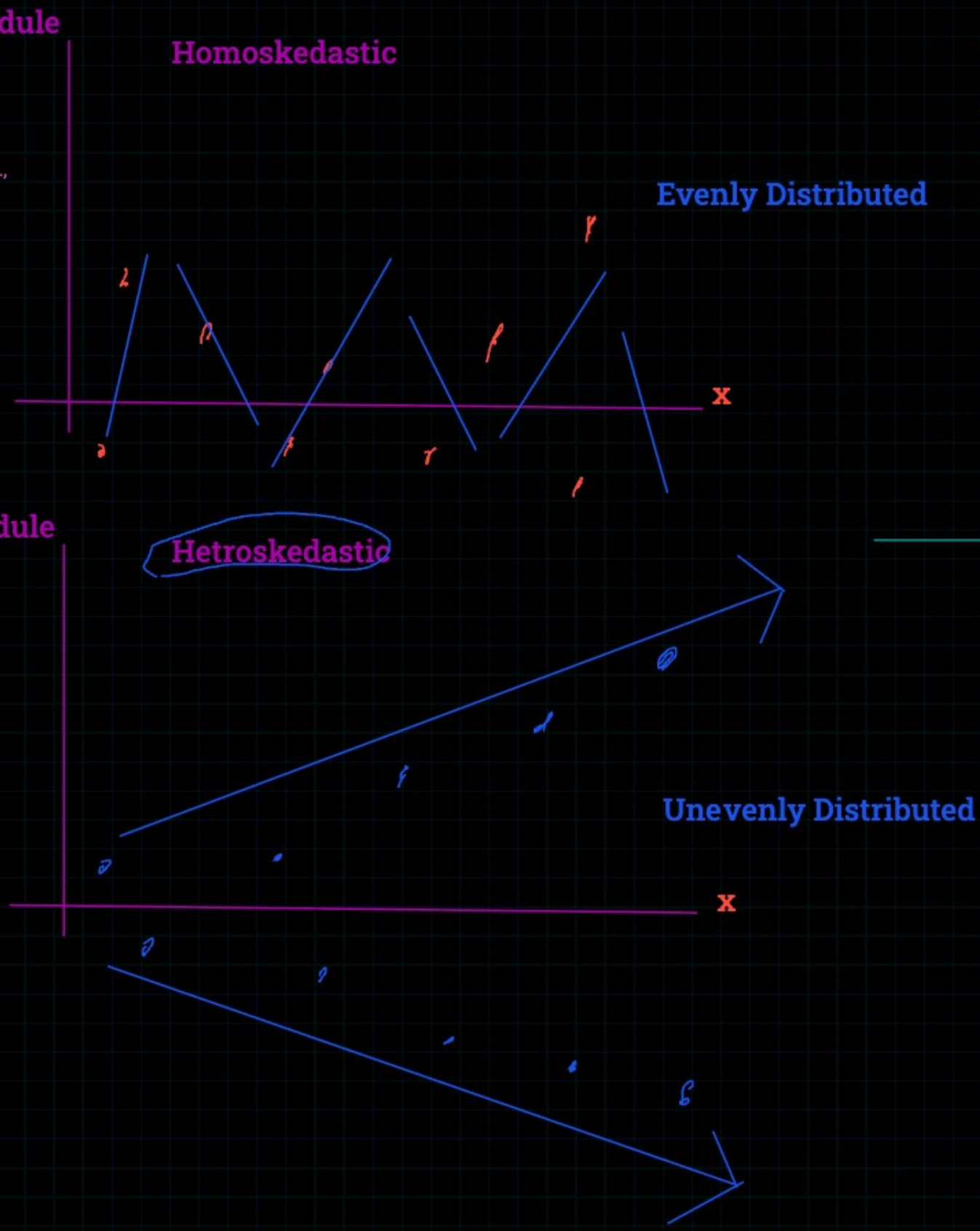
5. Homoskedasticity

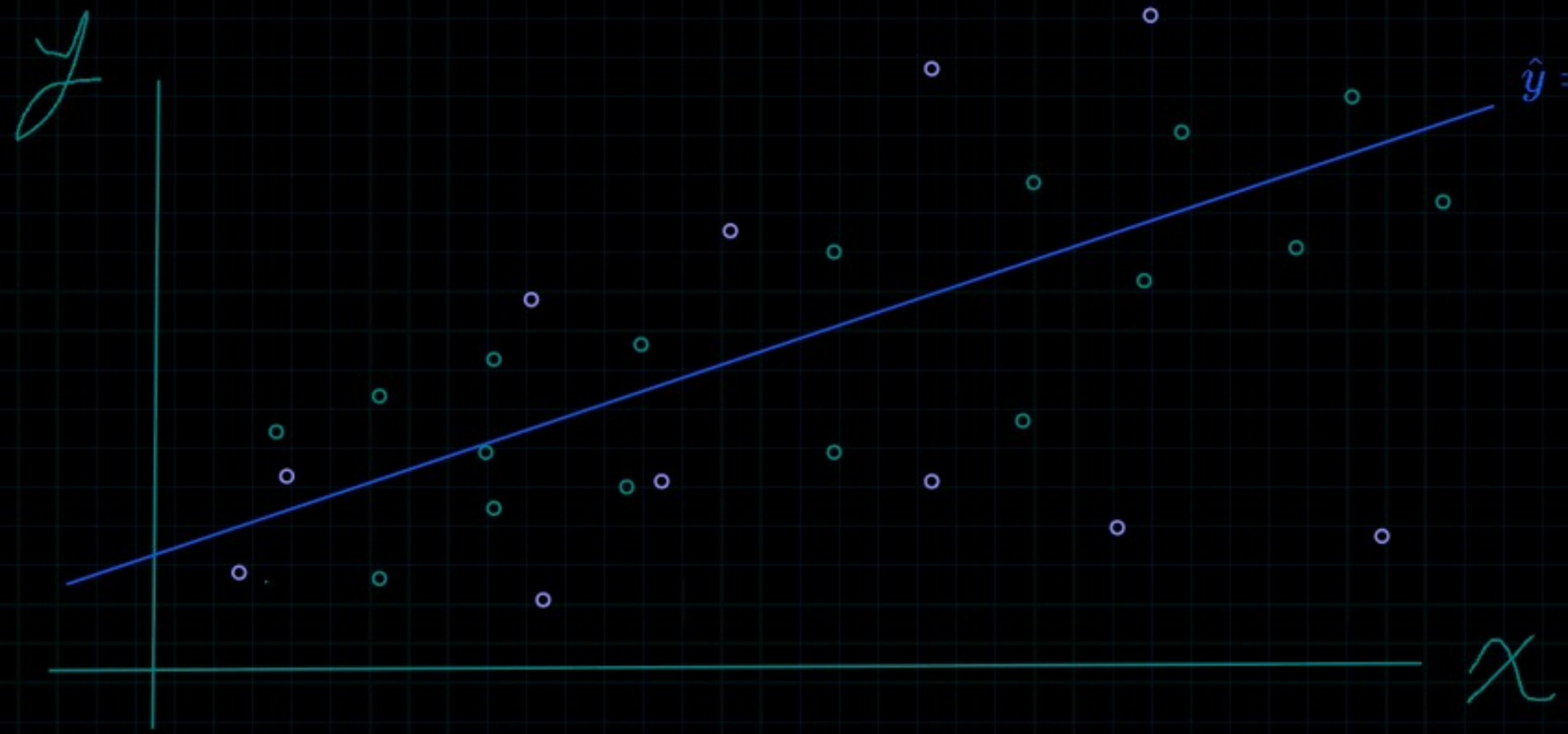
$\text{Var}(\mu | x_1, \ldots, x_k) = \sigma^2$ **where** σ^2 **is a constant.**

6. Normality: μ **is independent of** x **and is normally distributed with a mean** 0 **and variance** σ^2

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

In statistics, a residual is
the difference between an observed value and the value predicted by a regression model, i.e.,
$\hat{y} - y$





Then, for green indep
whereas for blue dots

The importance here is that we can then so form reasonable expectation from out line of best-fit as the variance of the population mean, given variables x_i does not increase nor decrease.

In a Linear Regression Model, we attempt to explain a dependent variable, y, using x, the independent variable.
Say we have k variables, then:
$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots \beta_k x_k$
What this means is that y depends on $x_1, x_2, \ldots, x_k$. Say for example,

om

Say we want to explain age:

$$\hat{age} = \beta_1 time$$

In fact this a causation (100% certainty) here

Say stock:

$$\hat{Stock} = \beta_0 + \beta_1(political - situation) + \beta_2(opinions)$$
$$+\beta_3(macroeconomic - factor) +, \ldots, +\beta_k(company's - financial - report)$$

Also interest rate, inflation rate ...

And so here we have a good explanation. BUT!! we would simplify the model!

Remarks. Think about it is it 'stronger' to explain
an event by 100 things or 3 things?

M:
In the case for stock 100 things is pretty strong but what about the overlaping
explanations, that wouldn't be concise. Amount in this sense does not matter.
What matters is the 'degree' to which the event is explained.

Comment: by assumption, we do not have perfect collinearity but overlapping explanations might exist.

Recall we have a parameter called R^2 : $\frac{SSE}{SST}$ i.e., the explanation for the variation around the mean
over the total variation.
Typically with higher R^2 we will be able to conclude 'greater degree' of explanation.
In particular, we can conduct two tests to test out model : (1) F test - for overall sigfinicance; (2) T-test - for individual parameters

With these we can then define $R^2_{adj} := 1 - \frac{SSR/(n-1-k)}{SST/n-1}$ which will be higher whenever we drop insignificant explanatory variable.