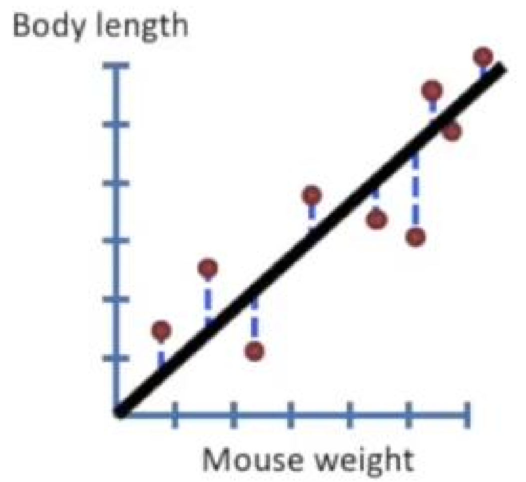
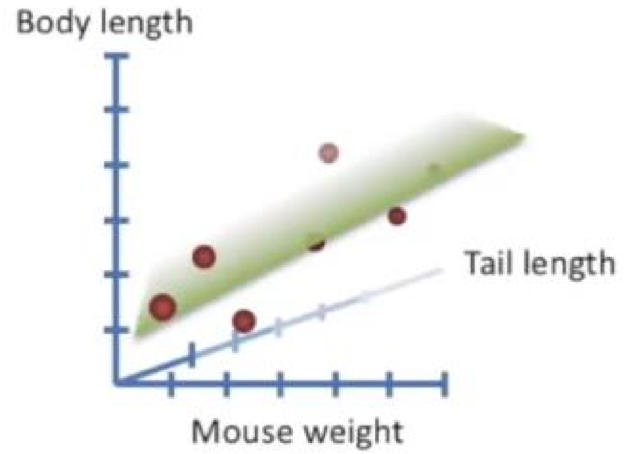
Simple regression



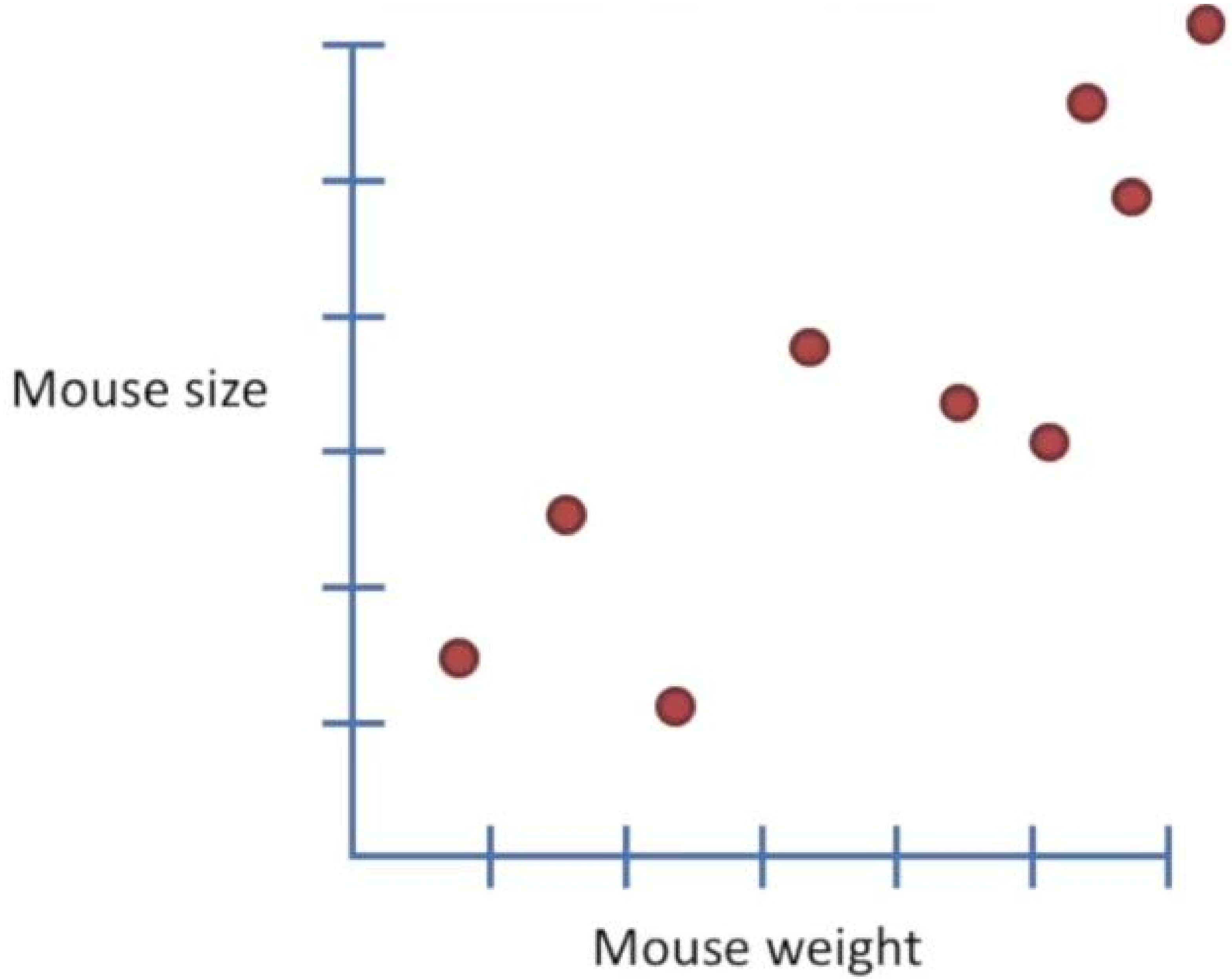
y = y-intercept + slope x

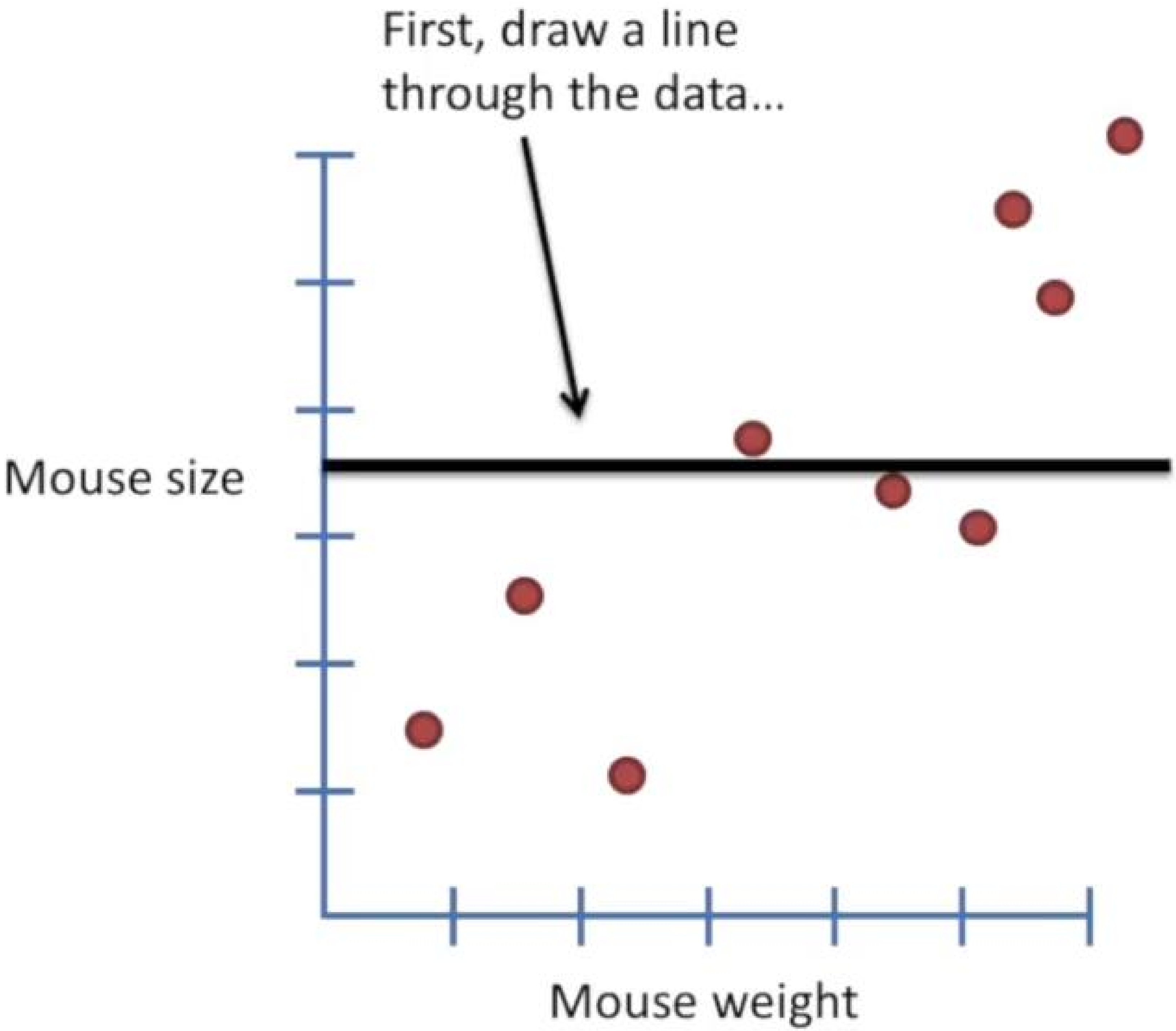
Calculating R^2 is the same for both simple and multiple regression

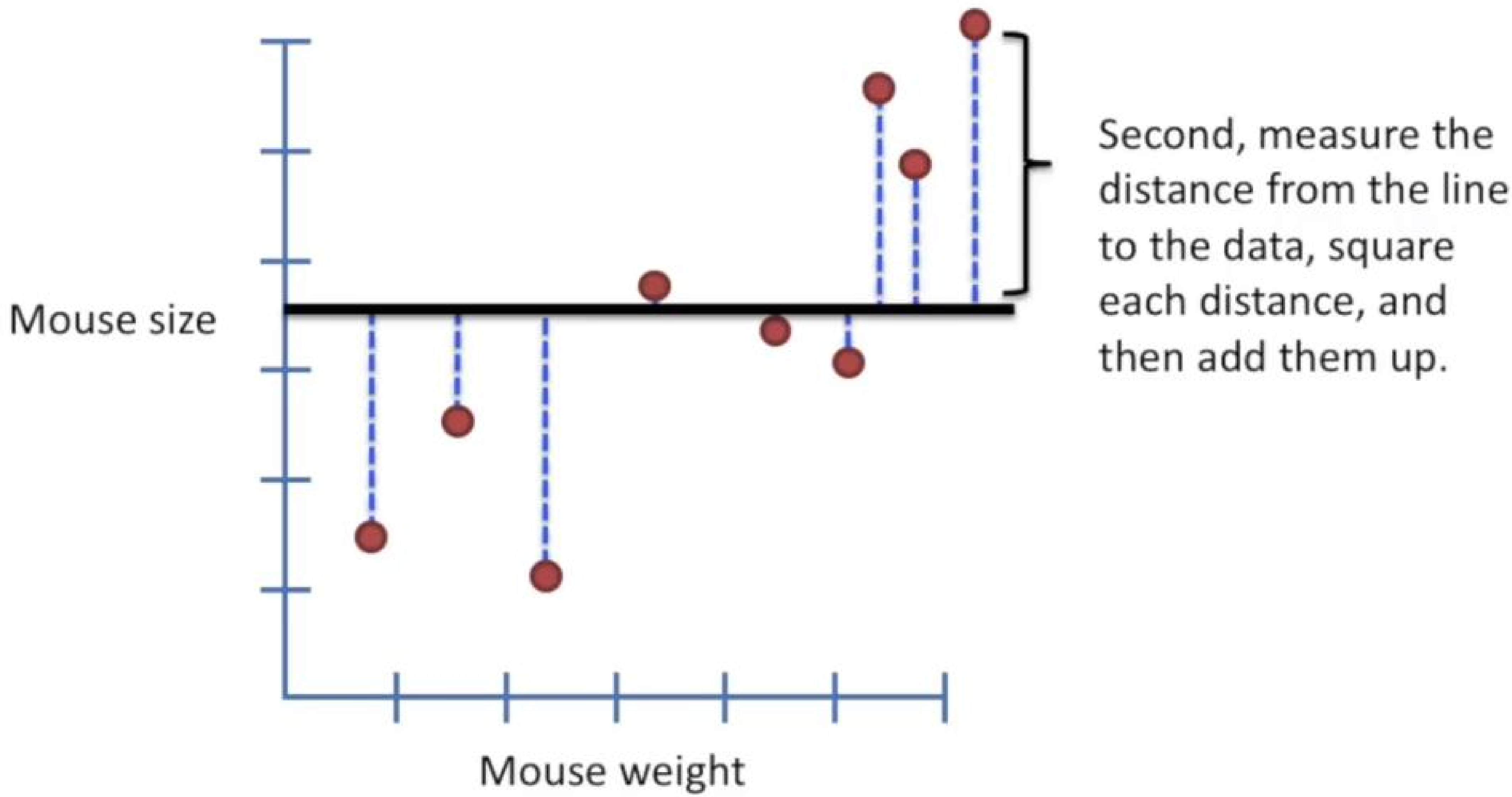
Multiple regression

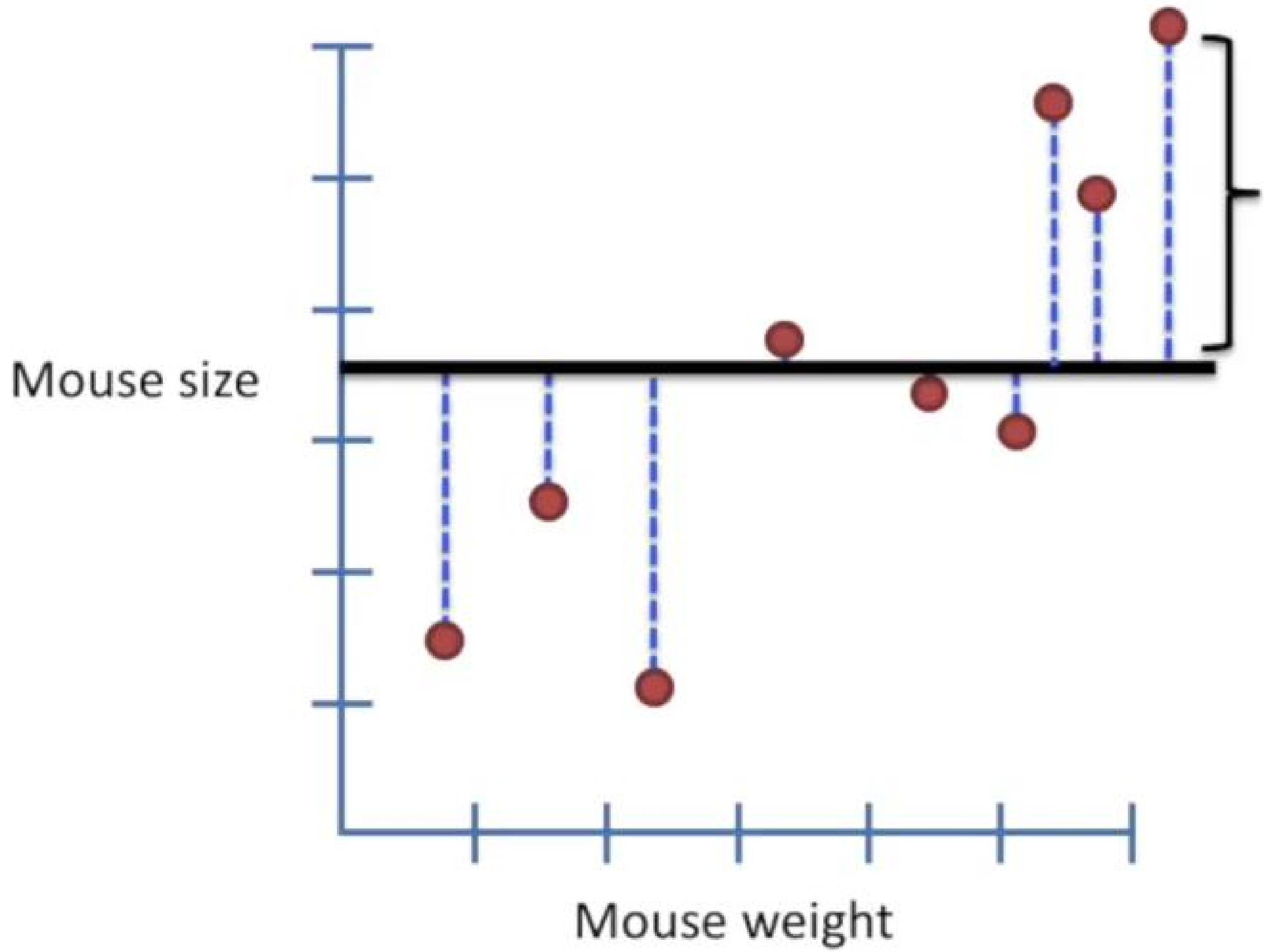


y = y-intercept + slope x + slope z


Mouse size
Mouse weight

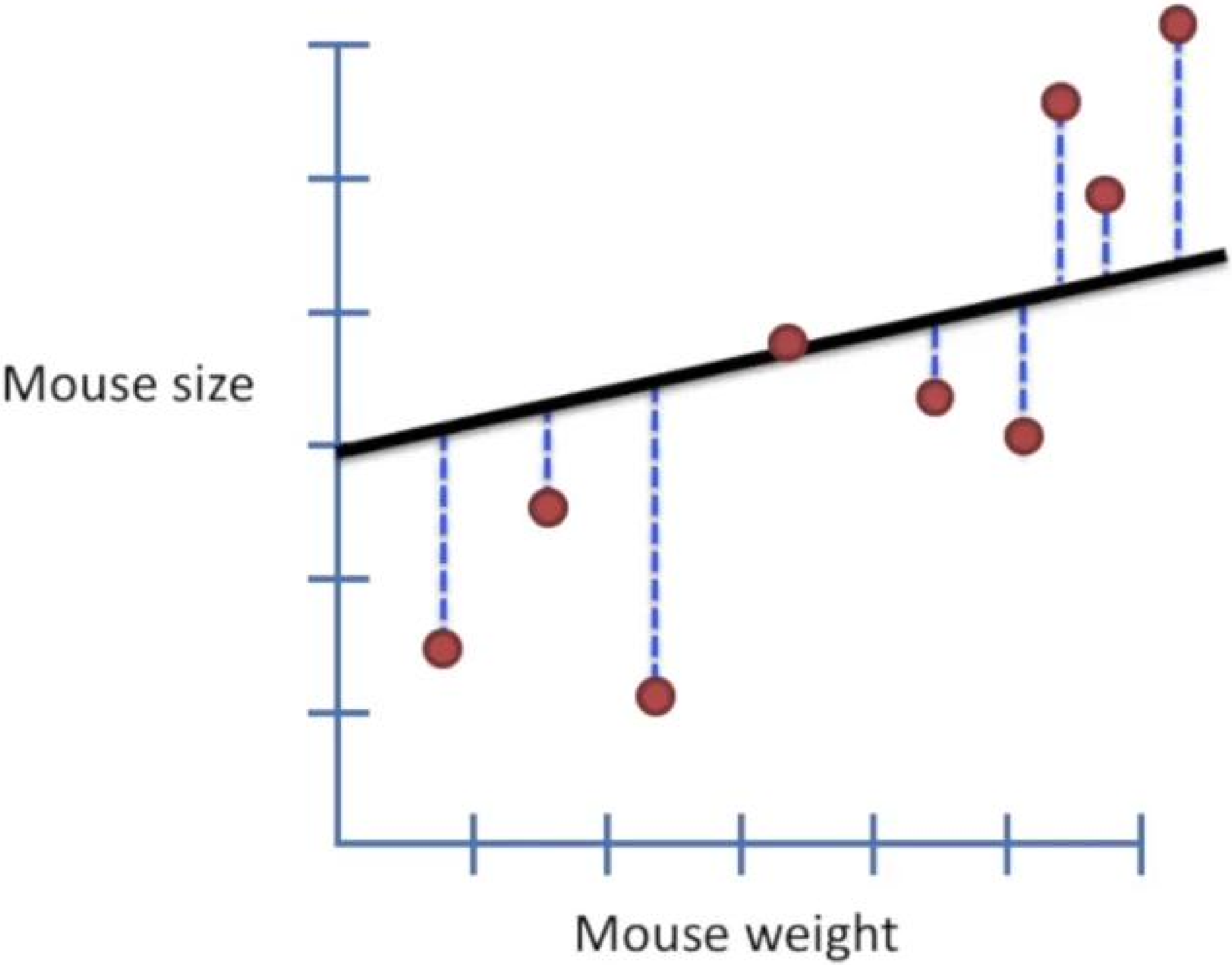

First, draw a line
through the data...
Mouse size
Mouse weight


Mouse size
Mouse weight
Second, measure the distance from the line to the data, square each distance, and then add them up.

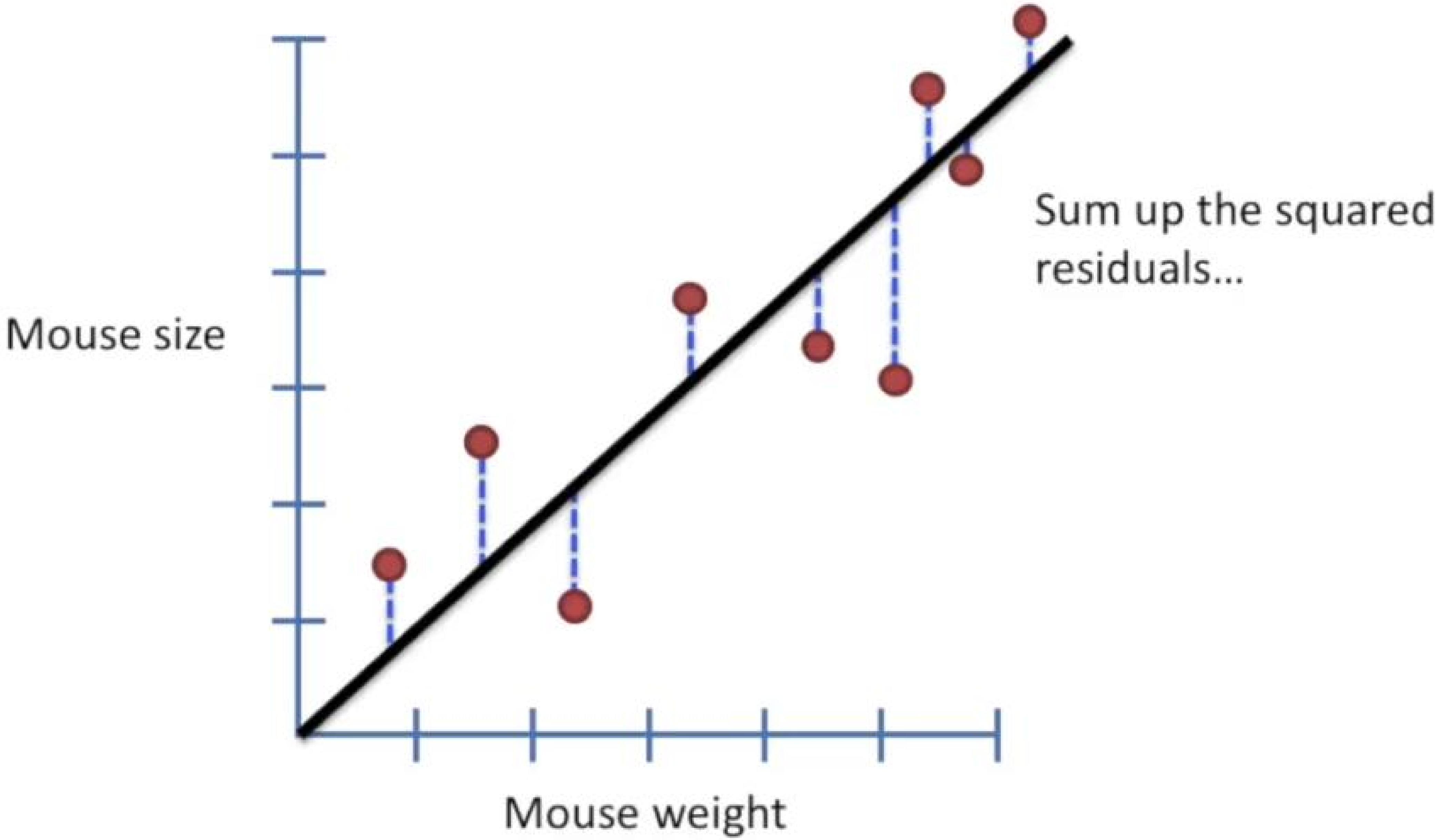


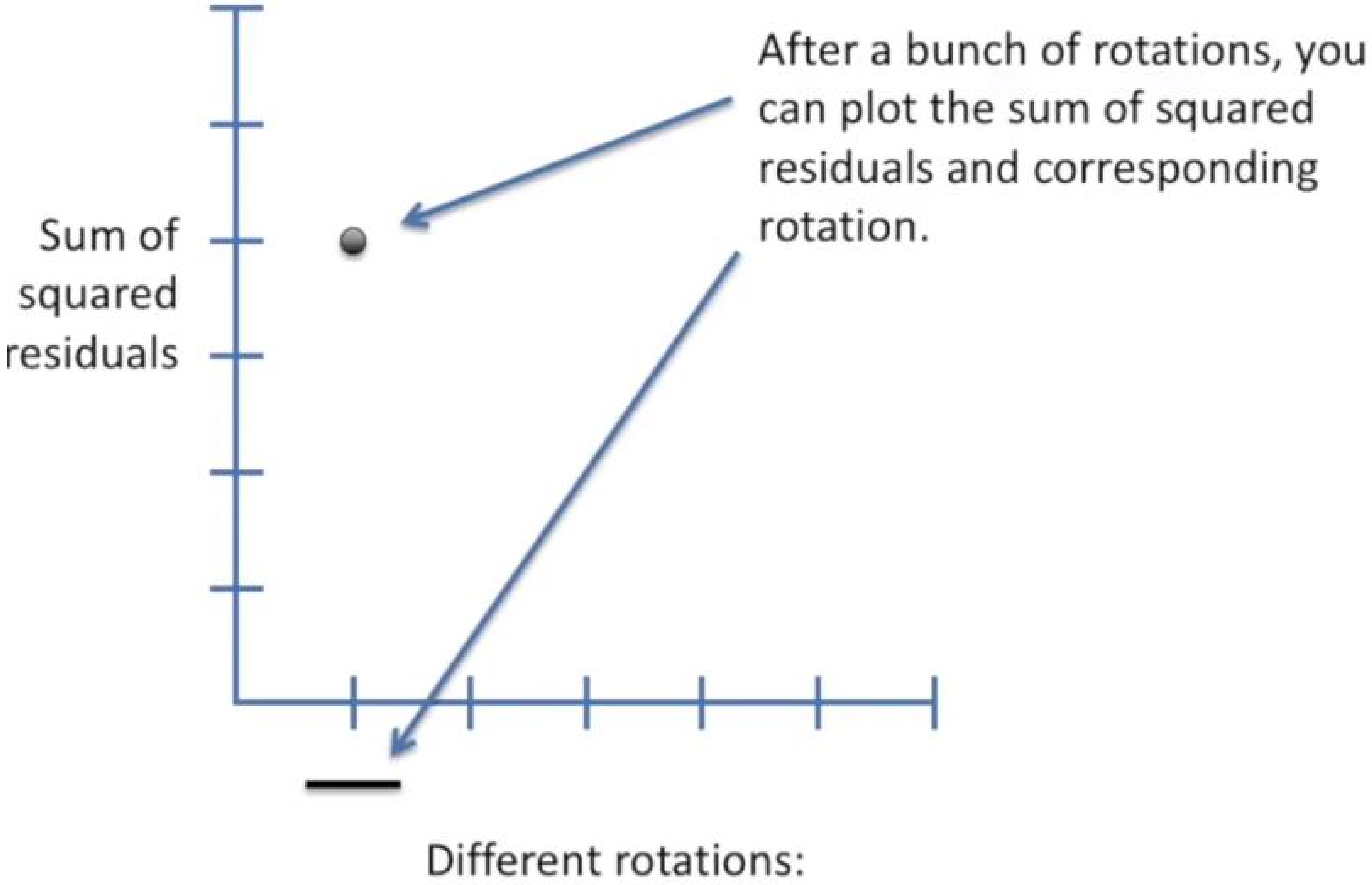
Second, measure the distance from the line to the data, square each distance, and then add them up.

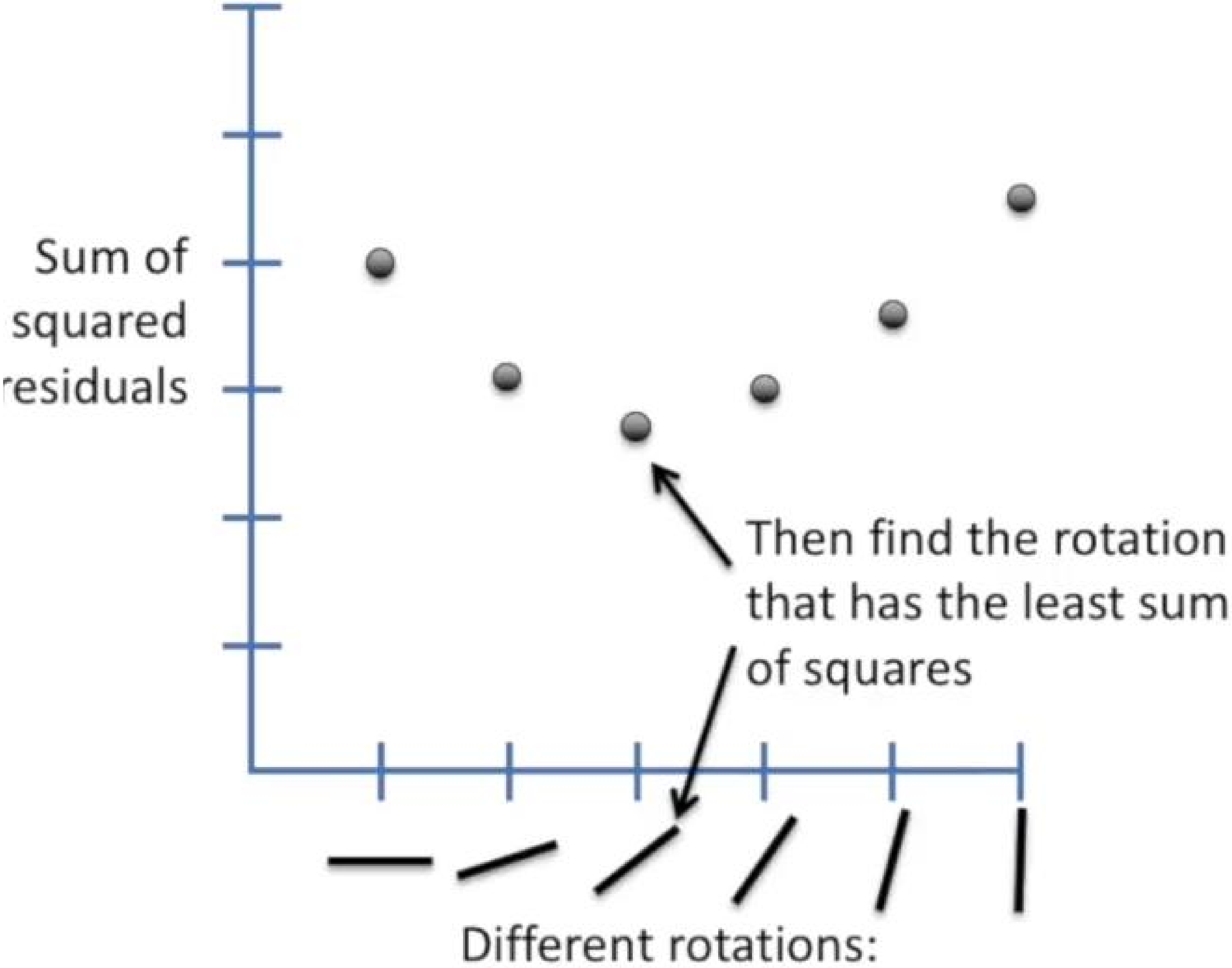
Terminology alert!
The distance from a line to a data point is called a "**residual**".

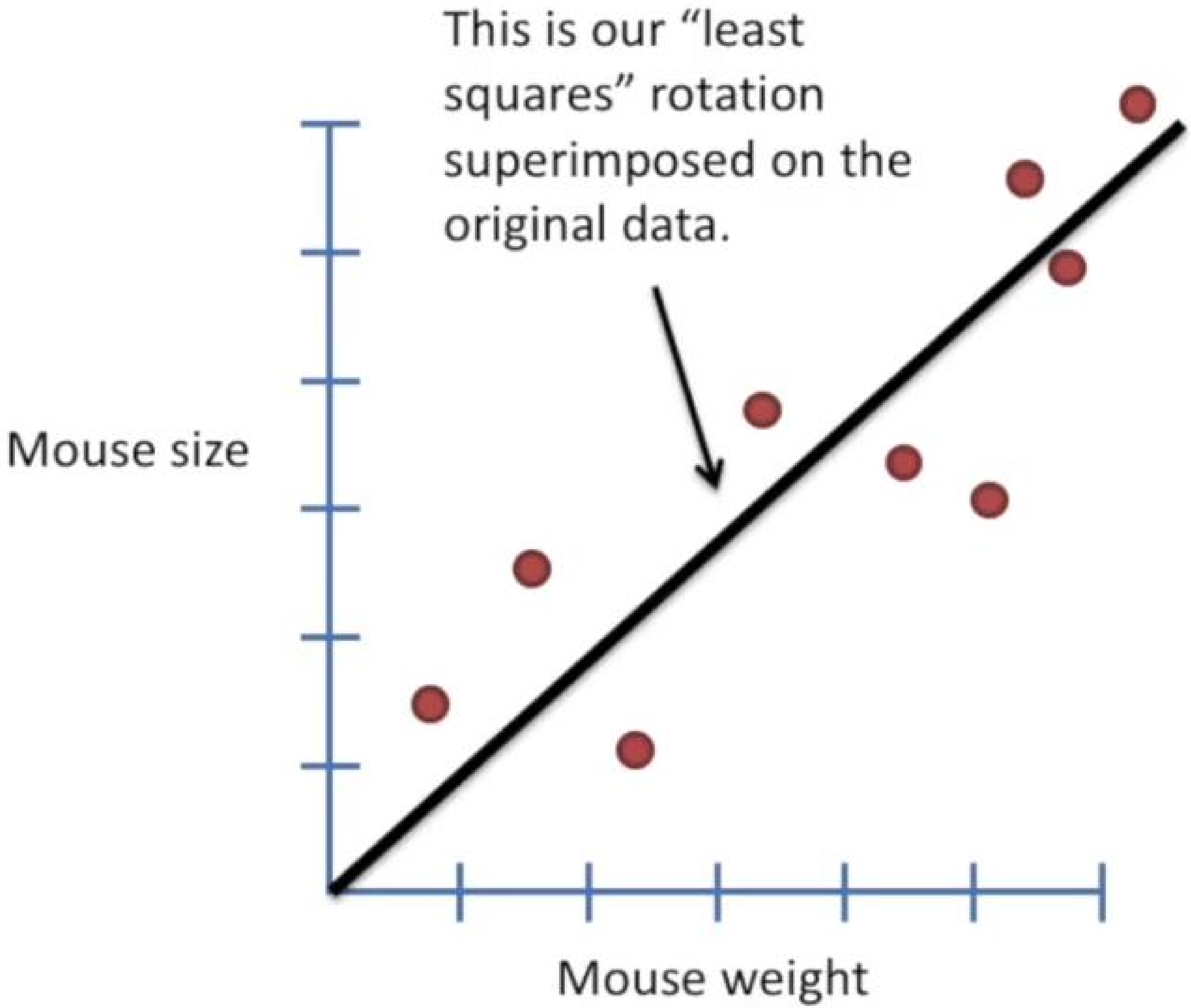


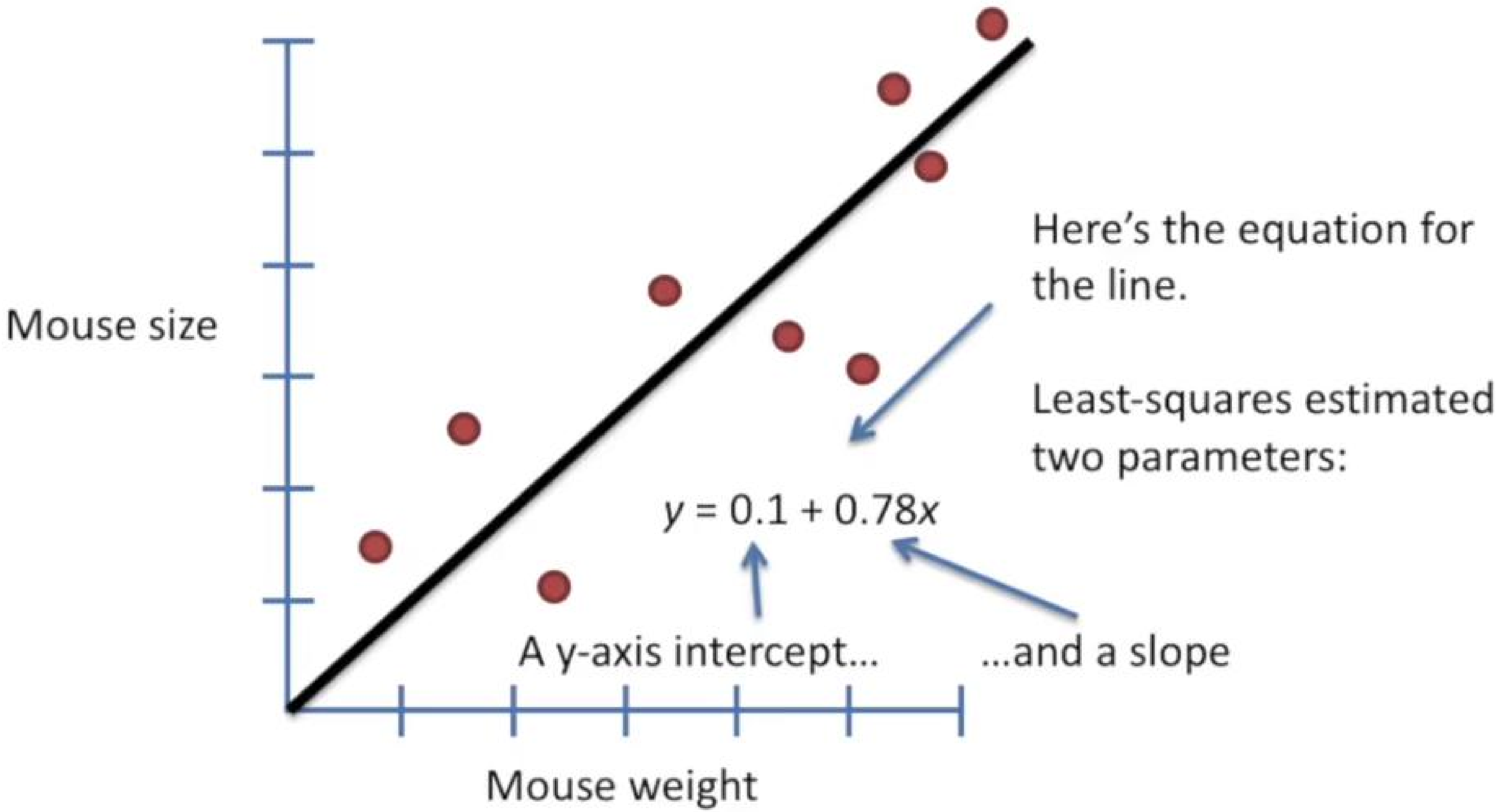
With the new line, measure the residuals, square them, and then sum up the squares.

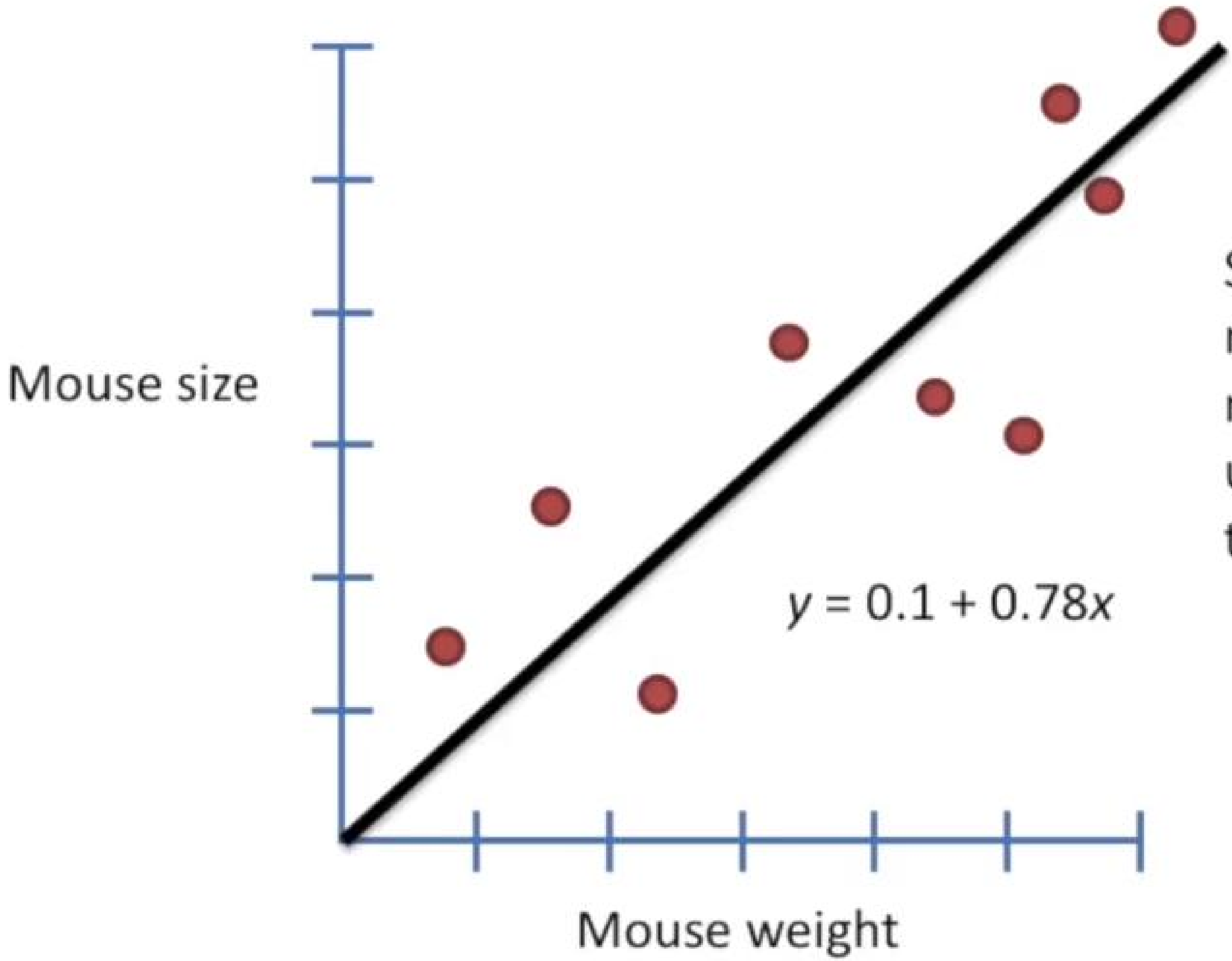

Sum up the squared residuals...
Mouse size
Mouse weight


After a bunch of rotations, you can plot the sum of squared residuals and corresponding rotation.
Sum of squared residuals
Different rotations:


Sum of
squared
residuals
Then find the rotation
that has the least sum
of squares
Different rotations:

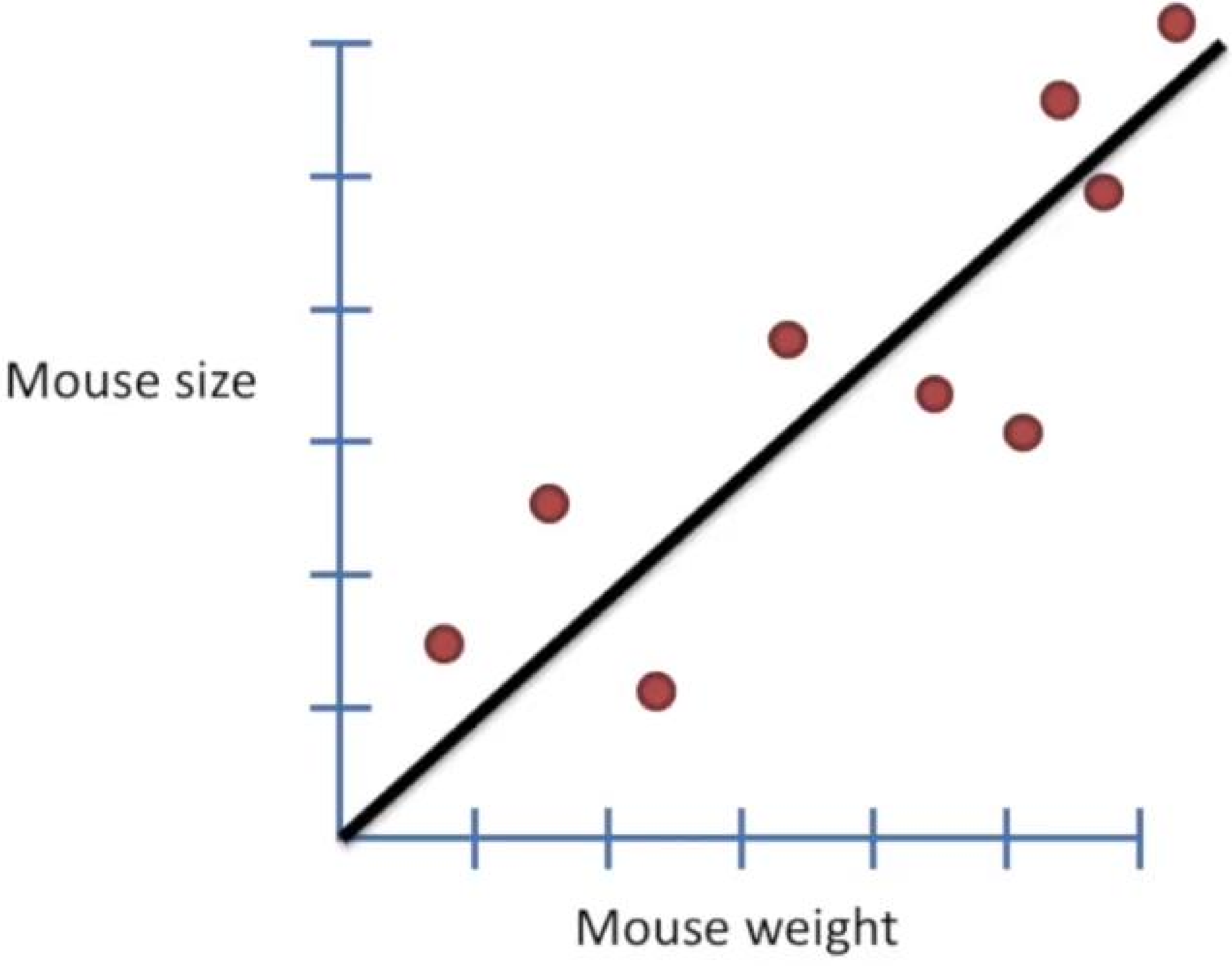

This is our "least squares" rotation superimposed on the original data.
Mouse size
Mouse weight


Here's the equation for the line.
Mouse size
Least-squares estimated two parameters:
y = 0.1 + 0.78x
A y-axis intercept...
...and a slope
Mouse weight

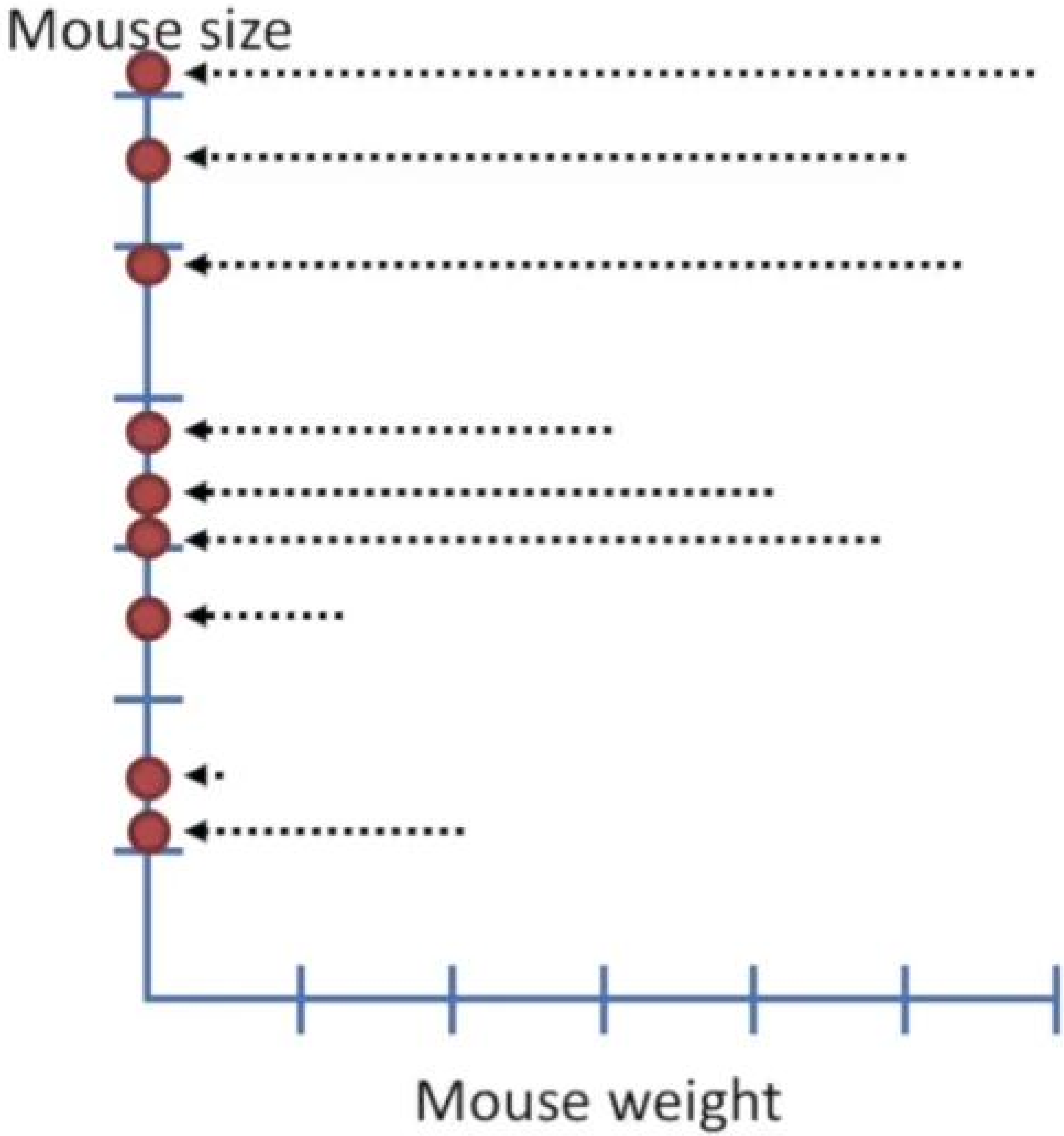


Since the slope is not 0, it means that knowing a mouse's weight will help us make a guess about that mouse's size.

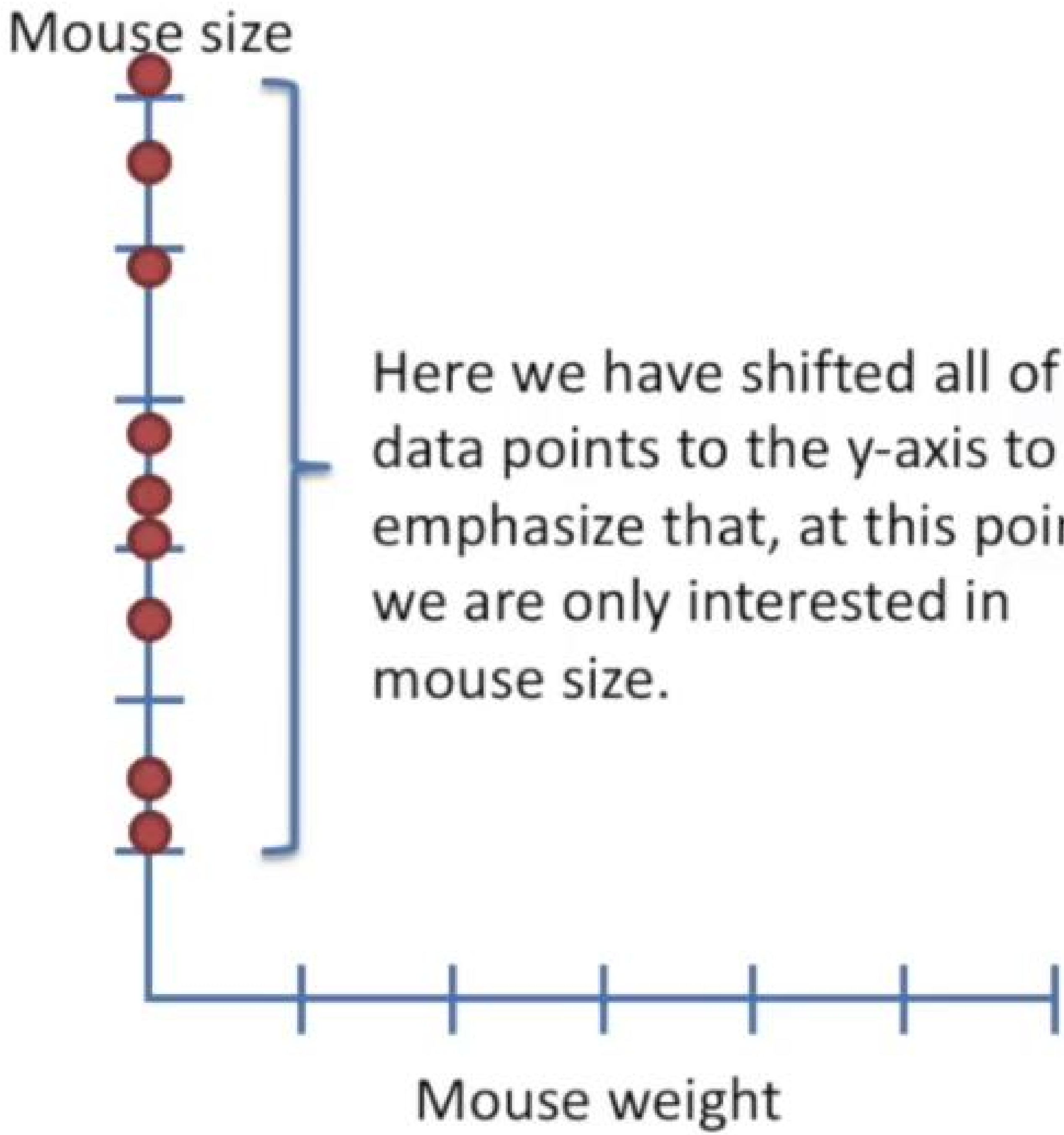
Calculating R^2 is the first step in determining how good that guess will be.



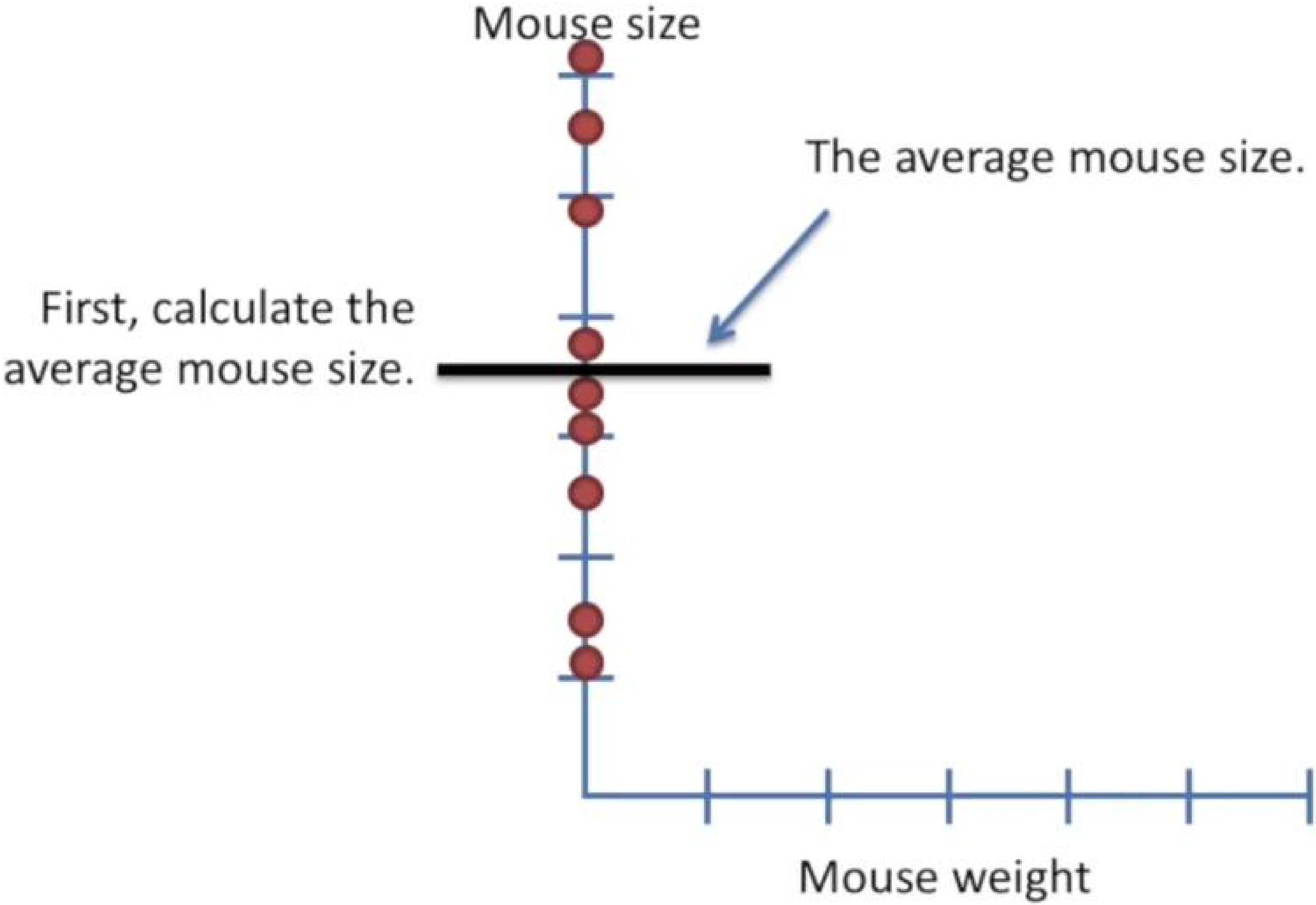
First, calculate the average mouse size.

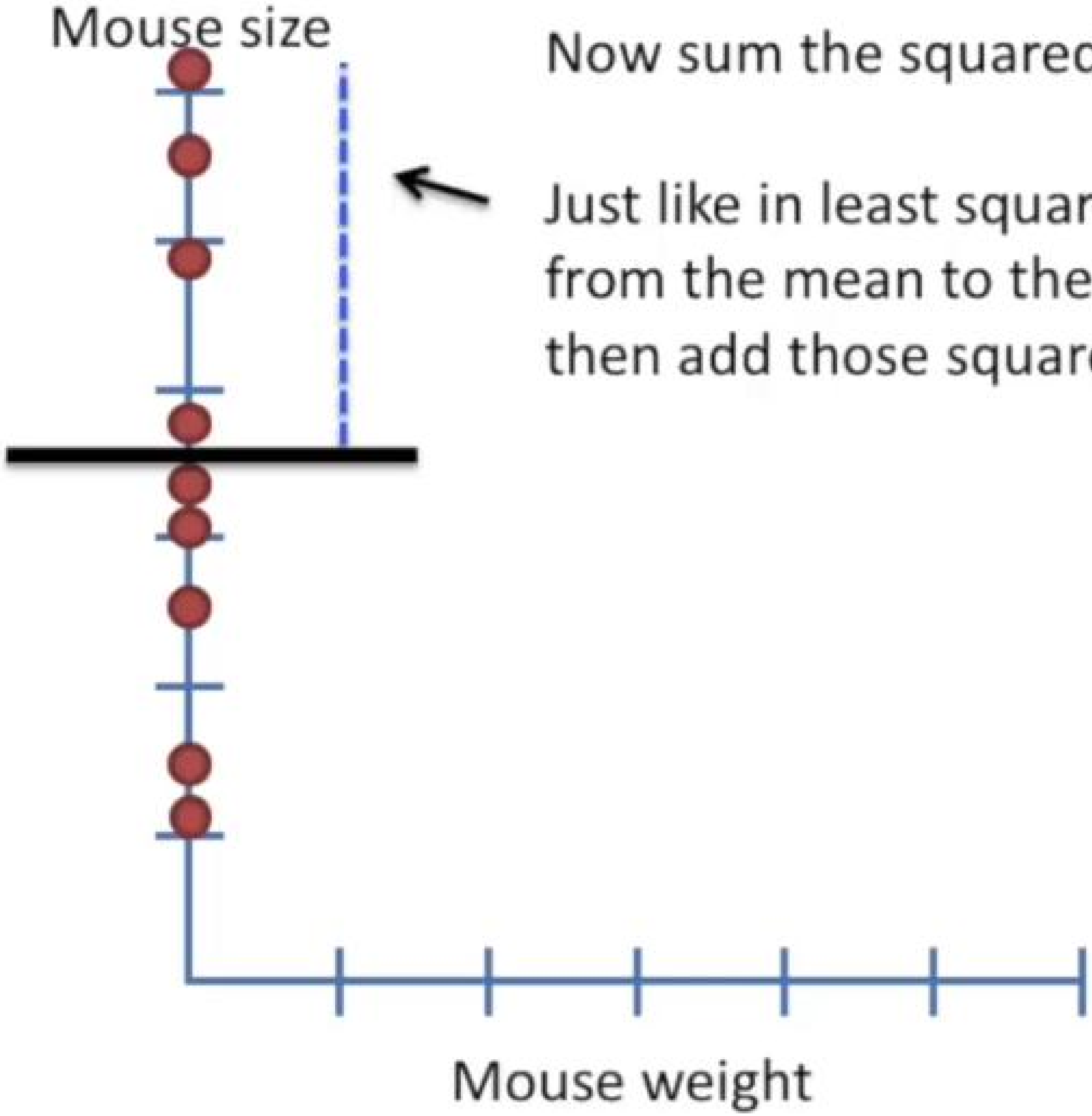


First, calculate the average mouse size.



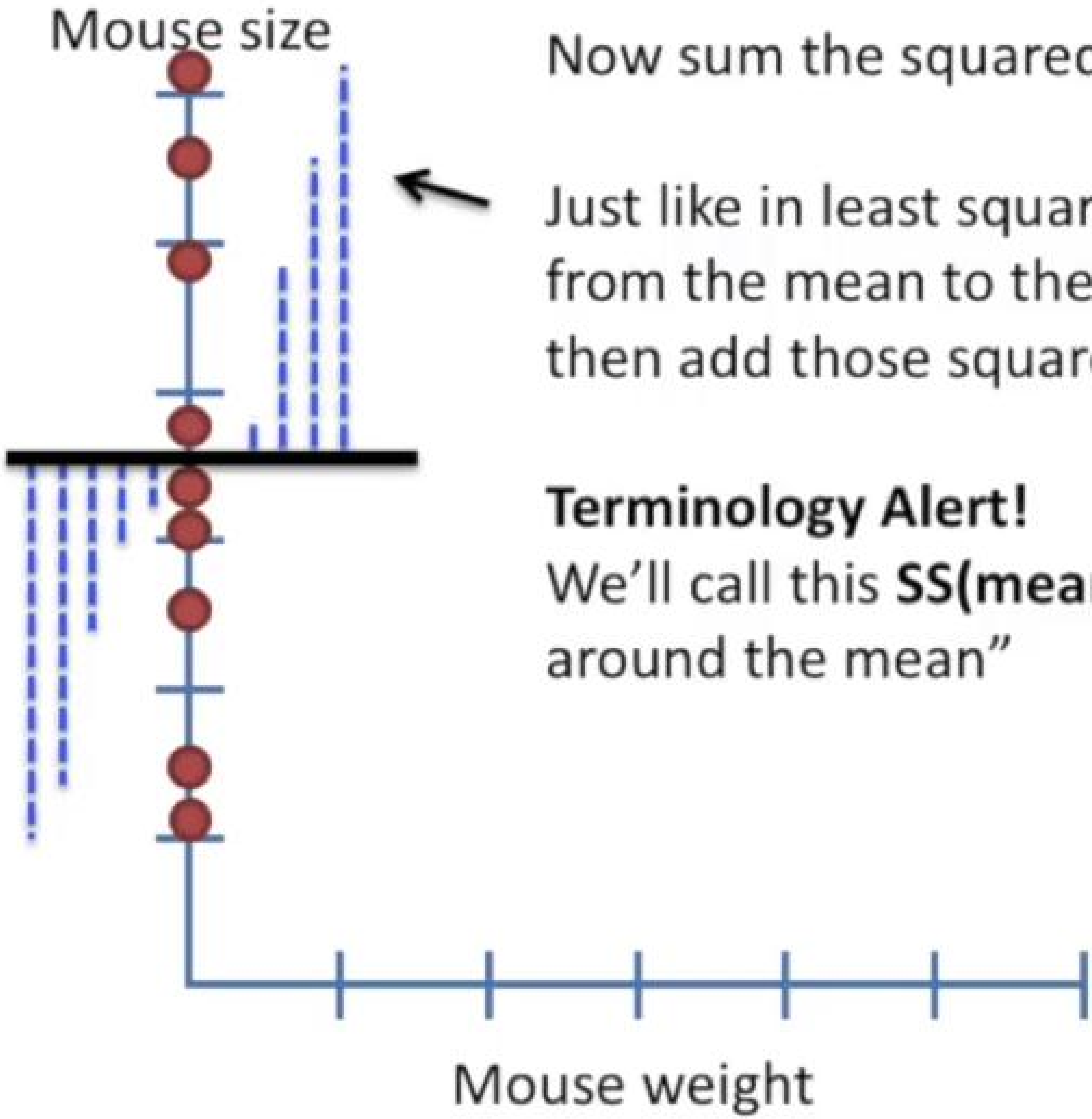
Here we have shifted all of the data points to the y-axis to emphasize that, at this point, we are only interested in mouse size.


Mouse size
The average mouse size.
First, calculate the
average mouse size.
Mouse weight



Now sum the squared residuals...

Just like in least squares, we measure the distance from the mean to the data point and square it, then add those squares together.

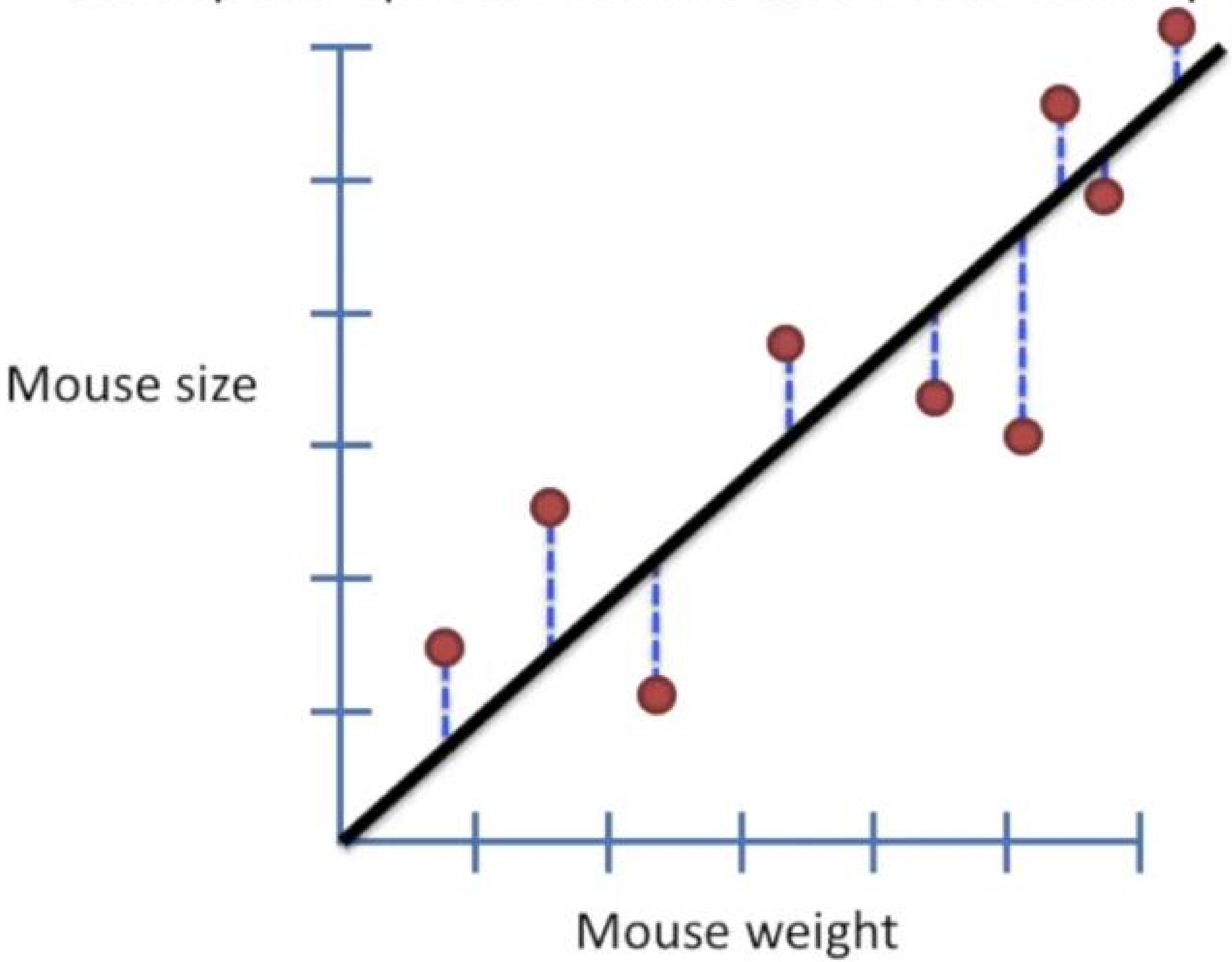


Now sum the squared residuals…

Just like in least squares, we measure the distance from the mean to the data point and square it, then add those squares together.

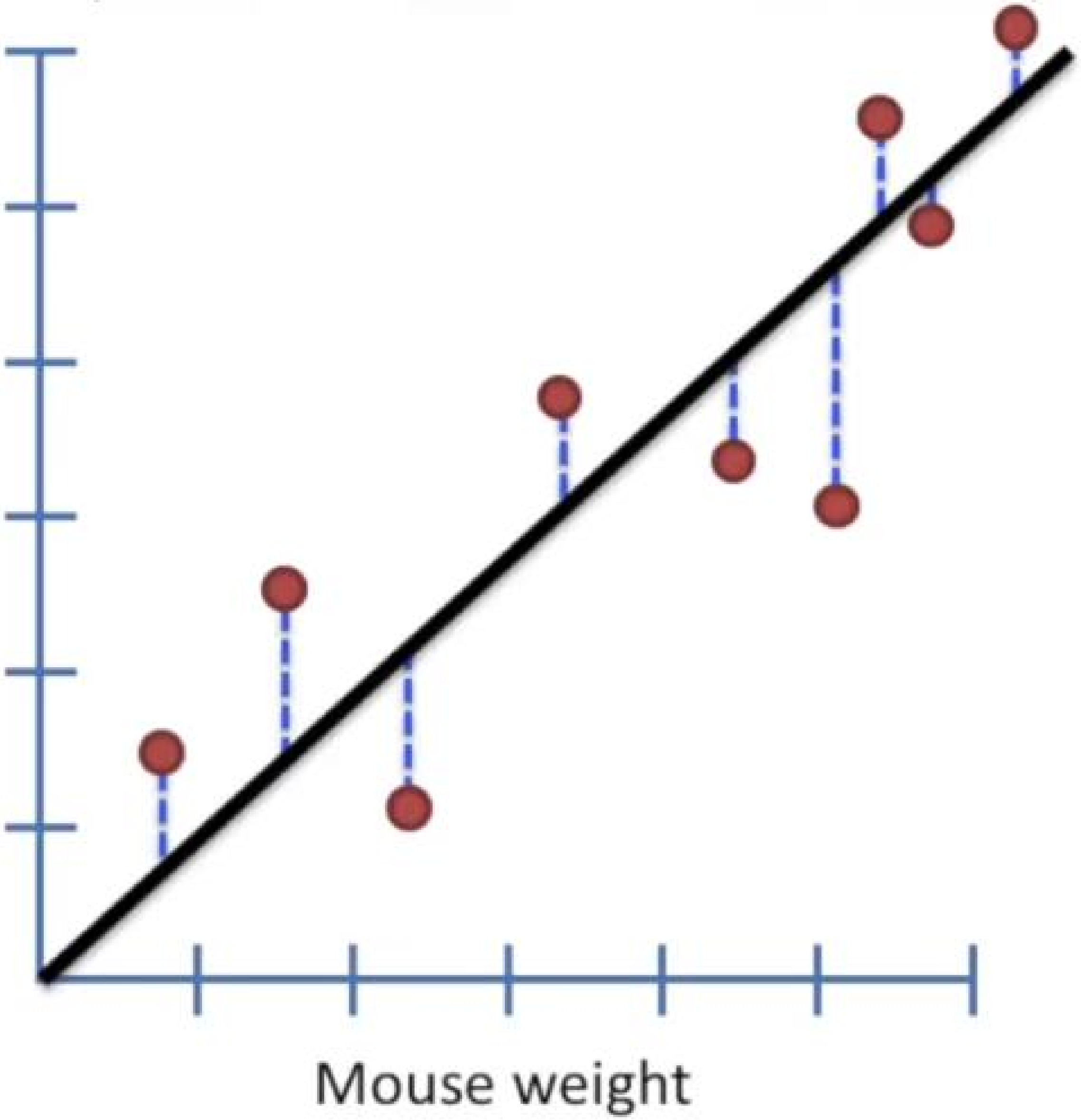
Terminology Alert!
We'll call this **SS(mean)**, for "sum of squares around the mean"

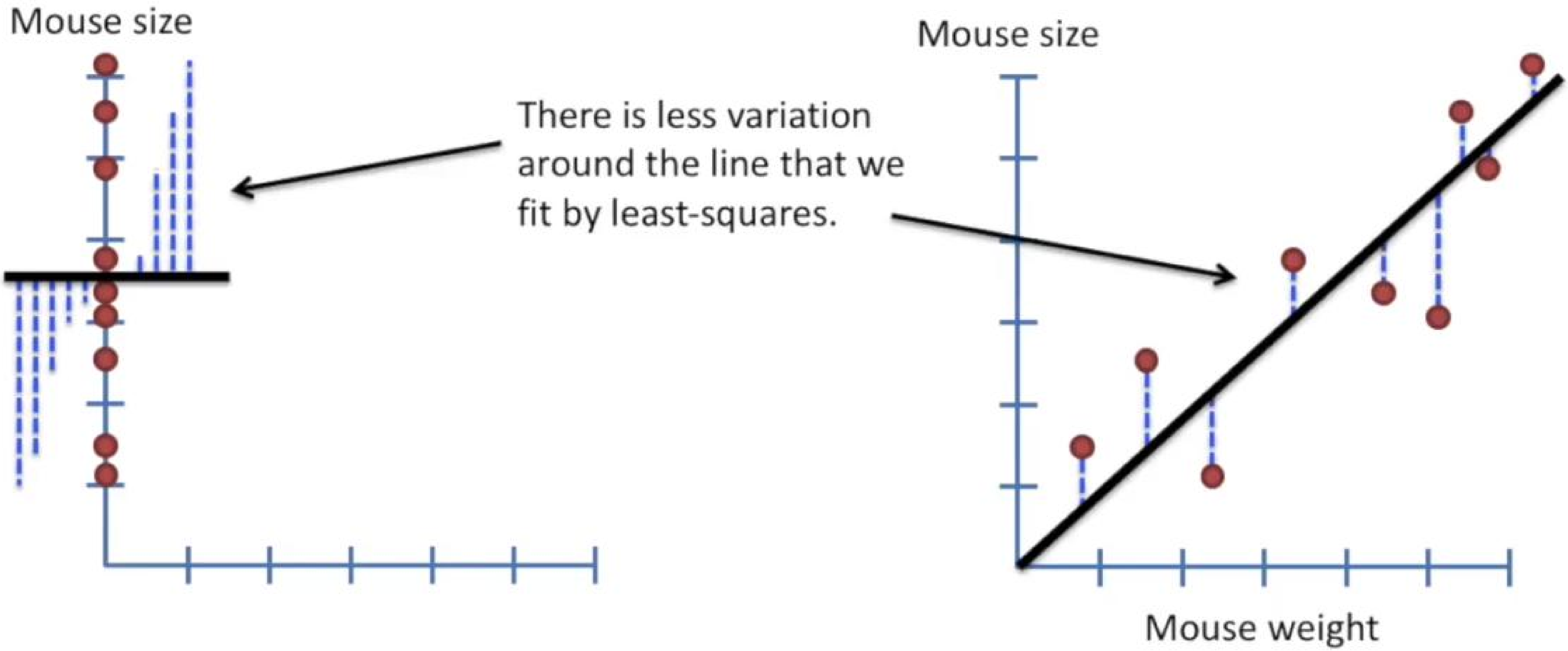
Now go back to the original plot.
Sum up the squared residuals around our least-squares fit.



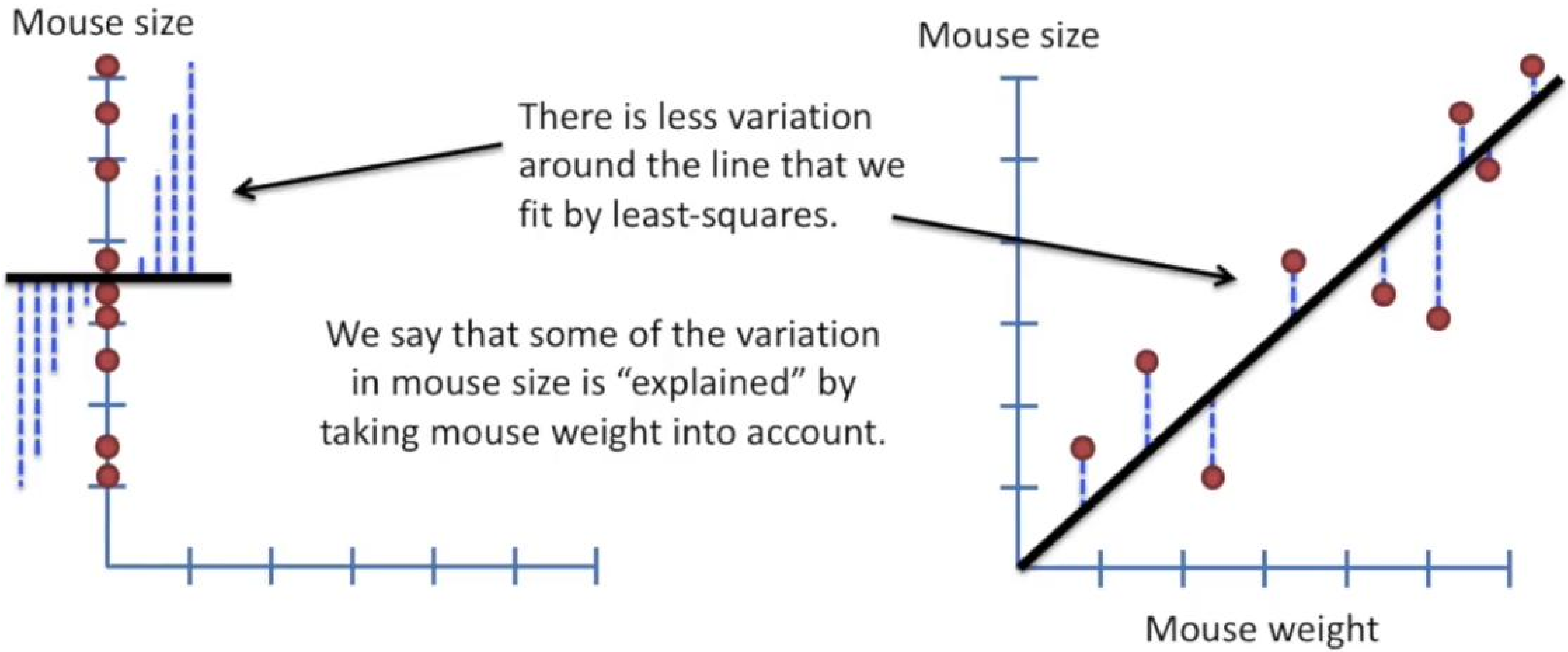
Now go back to the original plot.
Sum up the squared residuals around our least-squares fit.

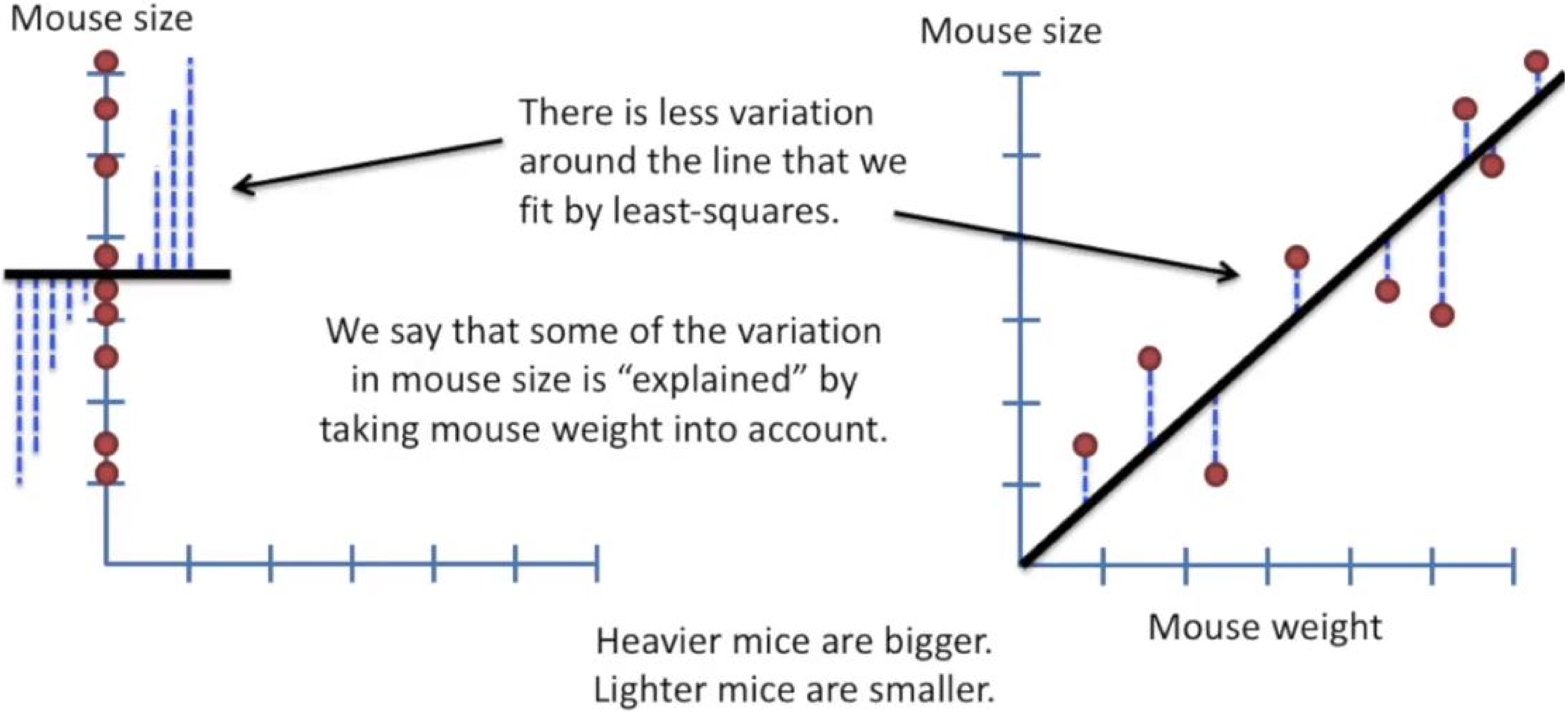
We'll call this **SS(fit)**, for the sum of squares around the least-squares fit.

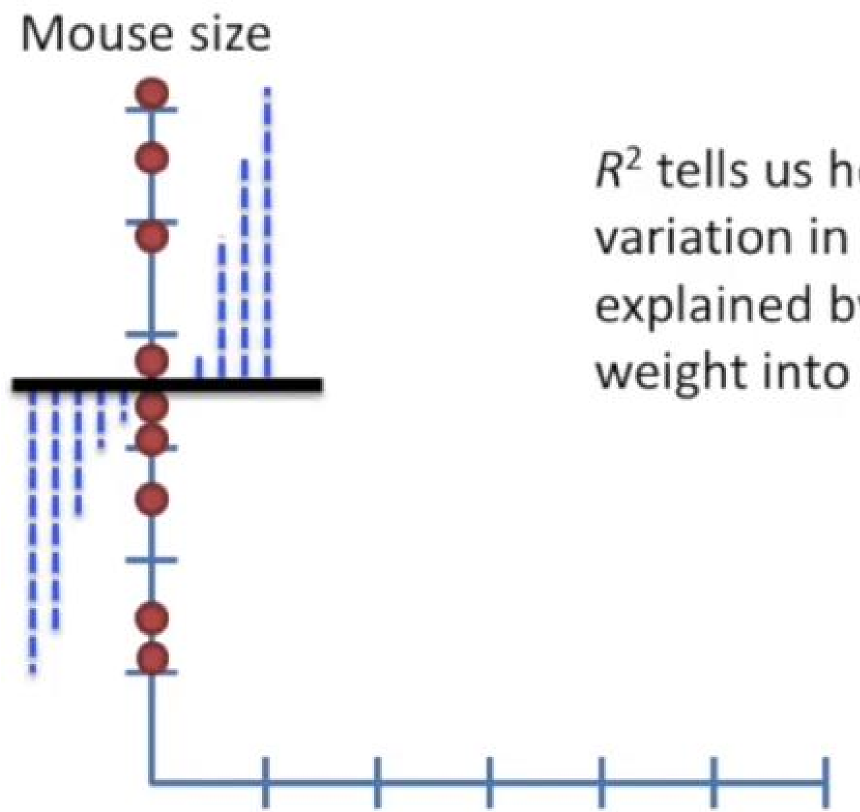




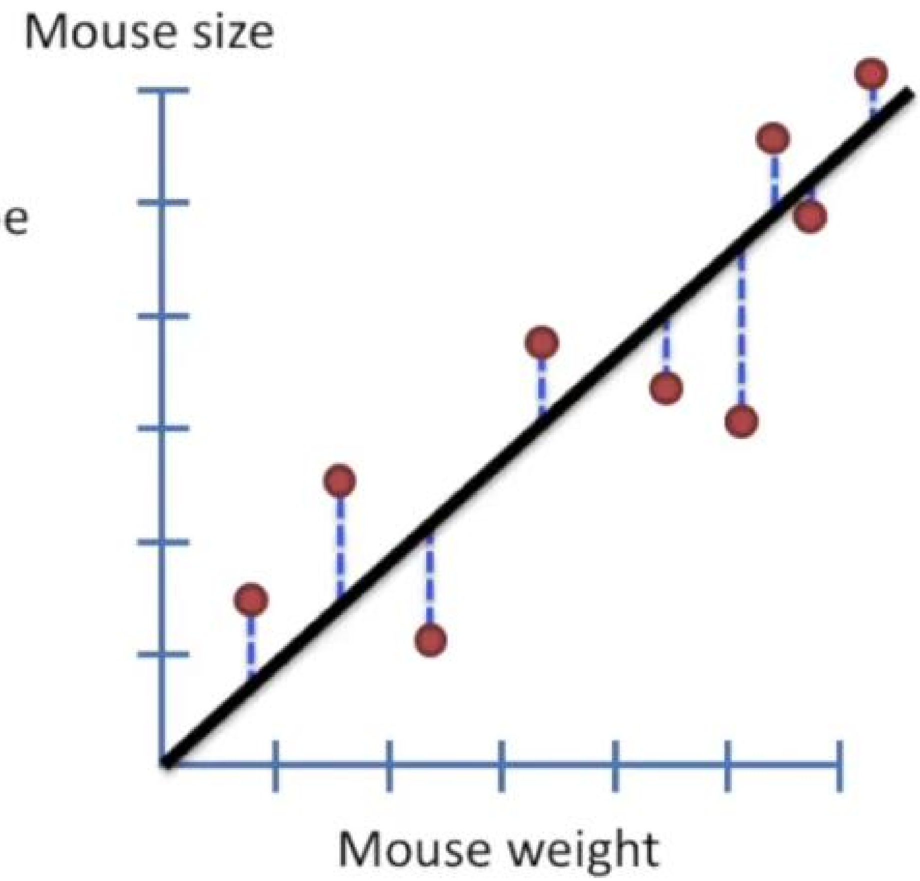
Mouse size
There is less variation around the line that we fit by least-squares.
Mouse size
Mouse weight

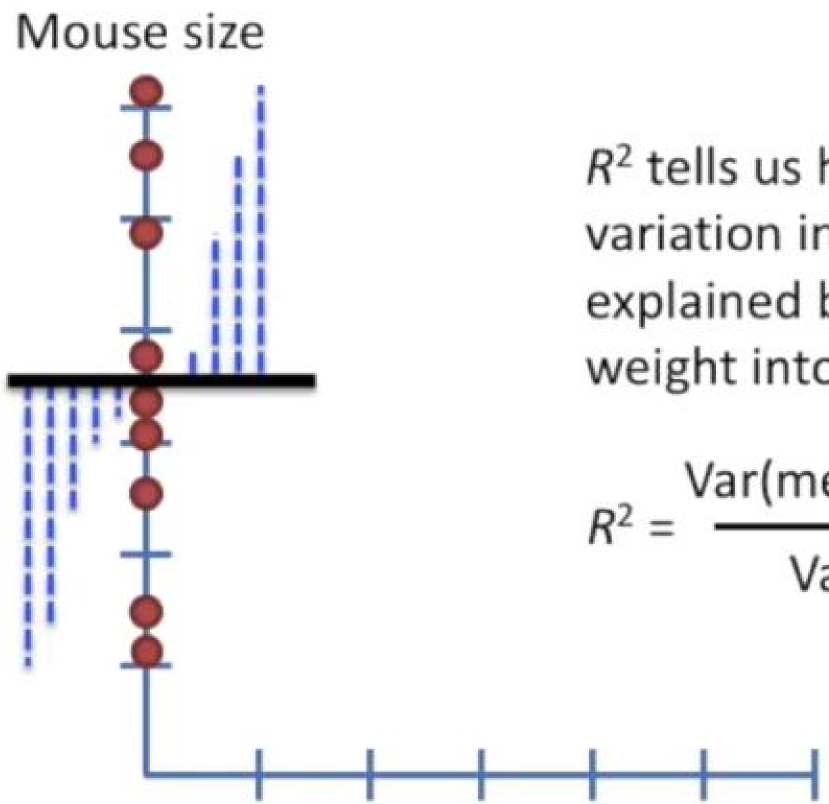

Mouse size
There is less variation
around the line that we
fit by least-squares.
We say that some of the variation
in mouse size is "explained" by
taking mouse weight into account.
Mouse size
Mouse weight


Mouse size
There is less variation around the line that we fit by least-squares.
We say that some of the variation in mouse size is "explained" by taking mouse weight into account.
Mouse size
Mouse weight
Heavier mice are bigger.
Lighter mice are smaller.



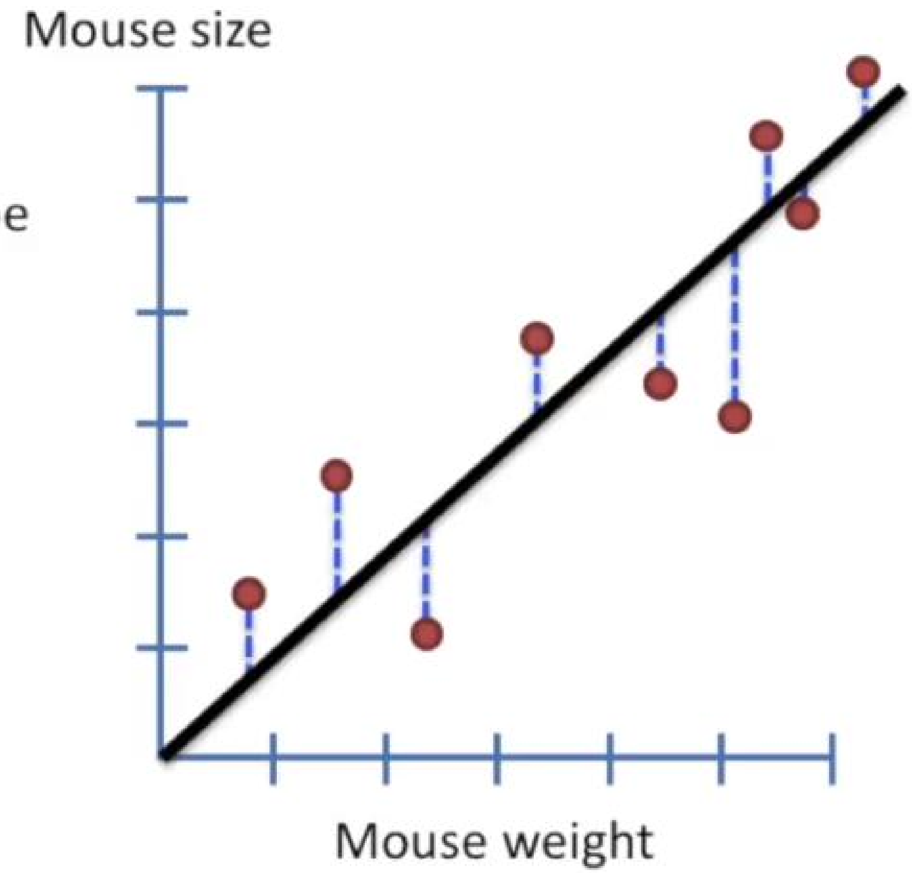
R^2 tells us how much of the variation in mouse size can be explained by taking mouse weight into account.



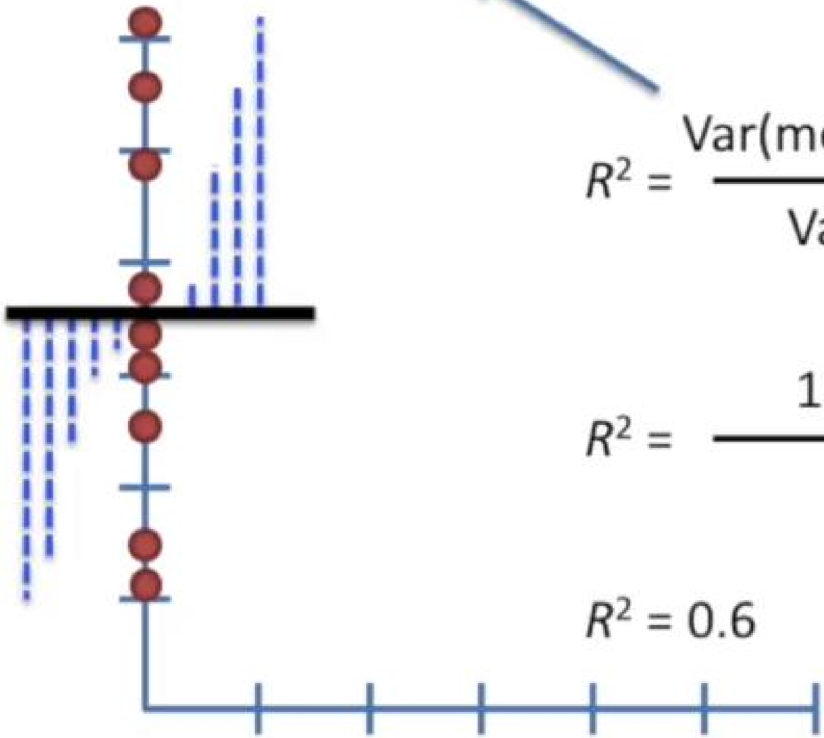


R^2 tells us how much of the variation in mouse size can be explained by taking mouse weight into account.

$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$



Var(mean) = 11.1

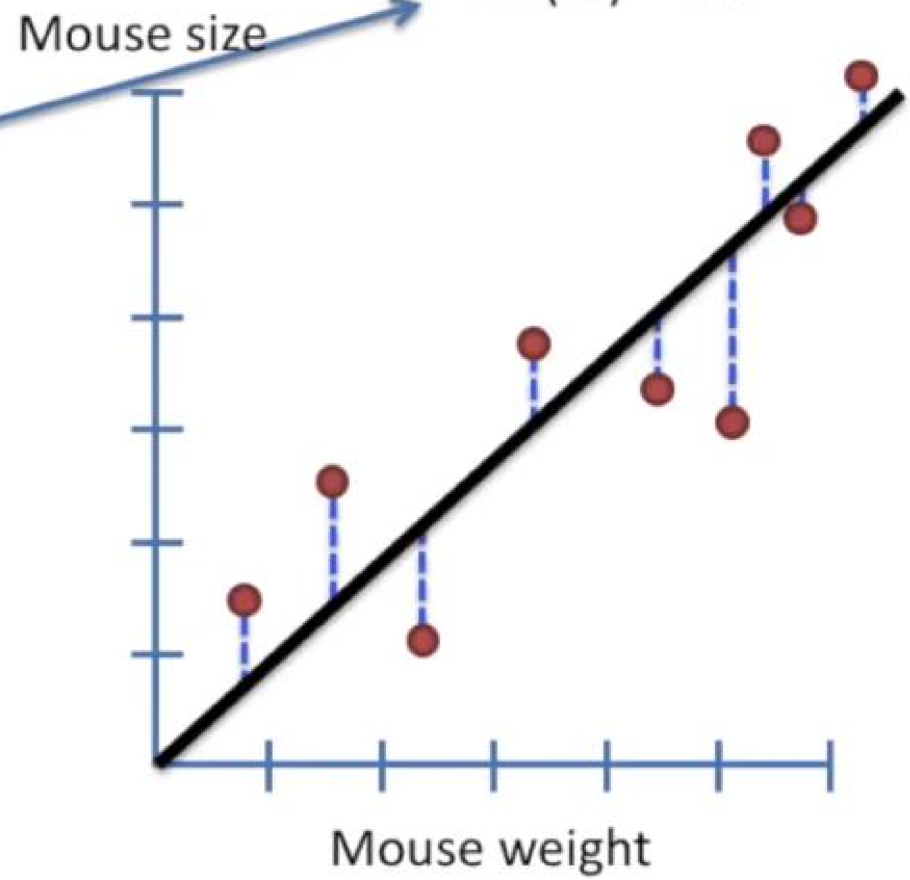


$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

$$R^2 = \frac{11.1 - 4.4}{11.1}$$

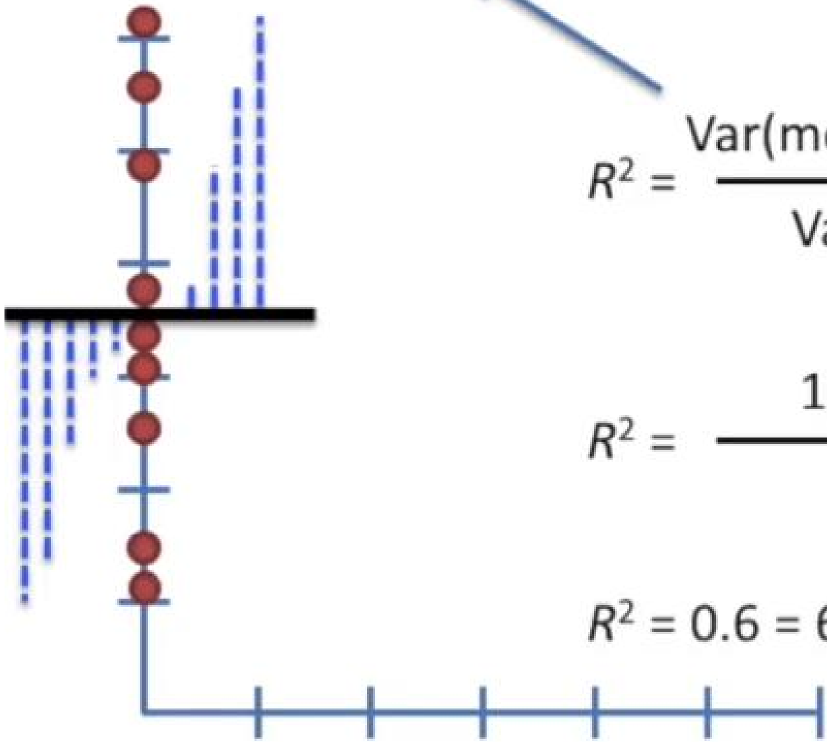
$$R^2 = 0.6$$

Var(fit) = 4.4



Var(mean) = 11.1

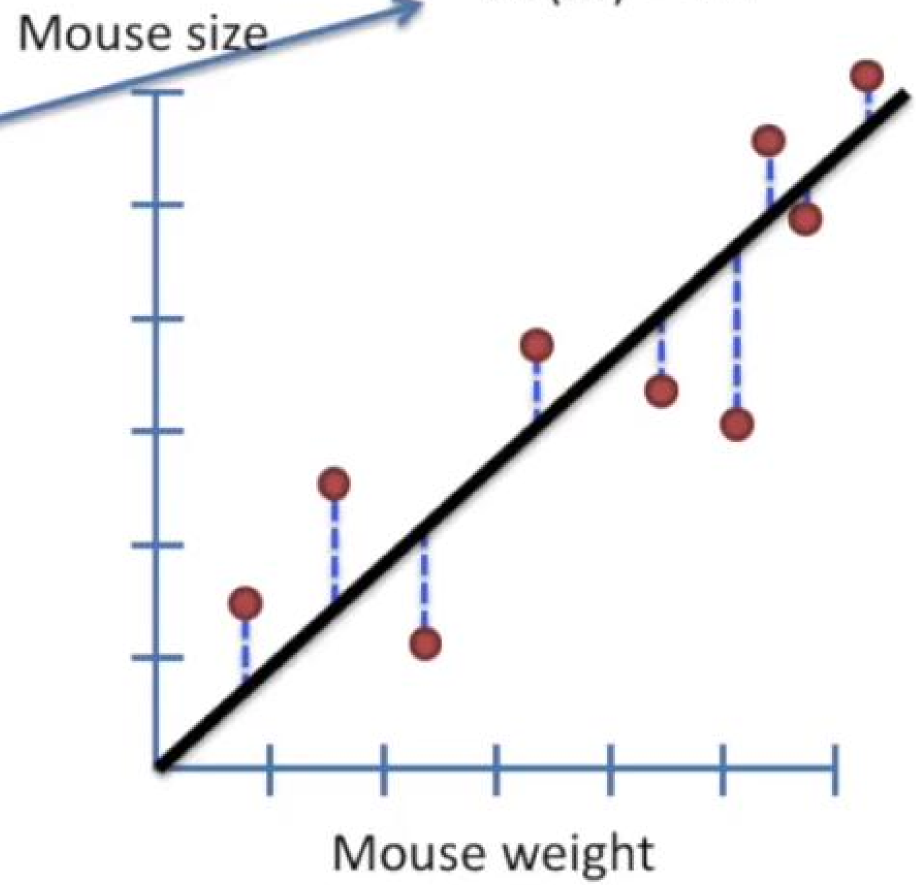
Var(fit) = 4.4



$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

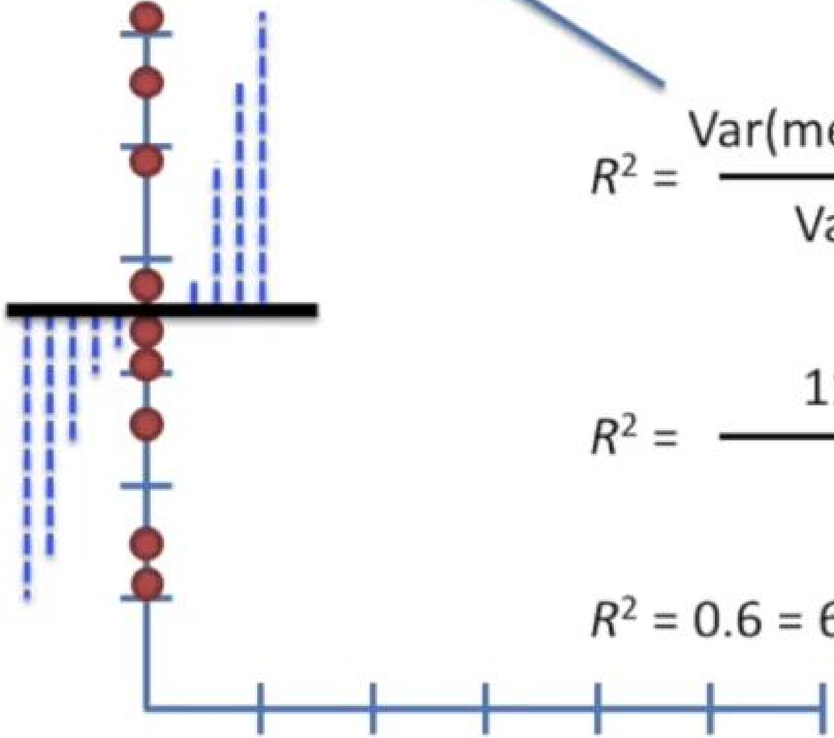
$$R^2 = \frac{11.1 - 4.4}{11.1}$$

$$R^2 = 0.6 = 60\%$$



Var(mean) = 11.1

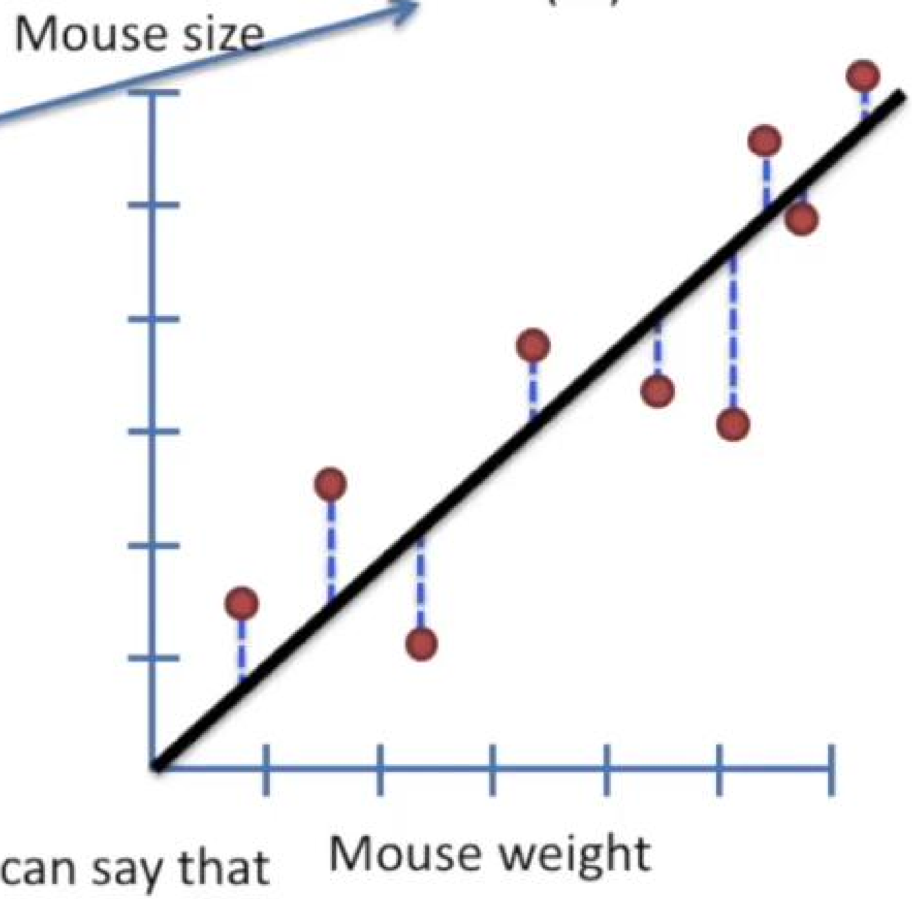
Var(fit) = 4.4



$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

$$R^2 = \frac{11.1 - 4.4}{11.1}$$

$$R^2 = 0.6 = 60\%$$

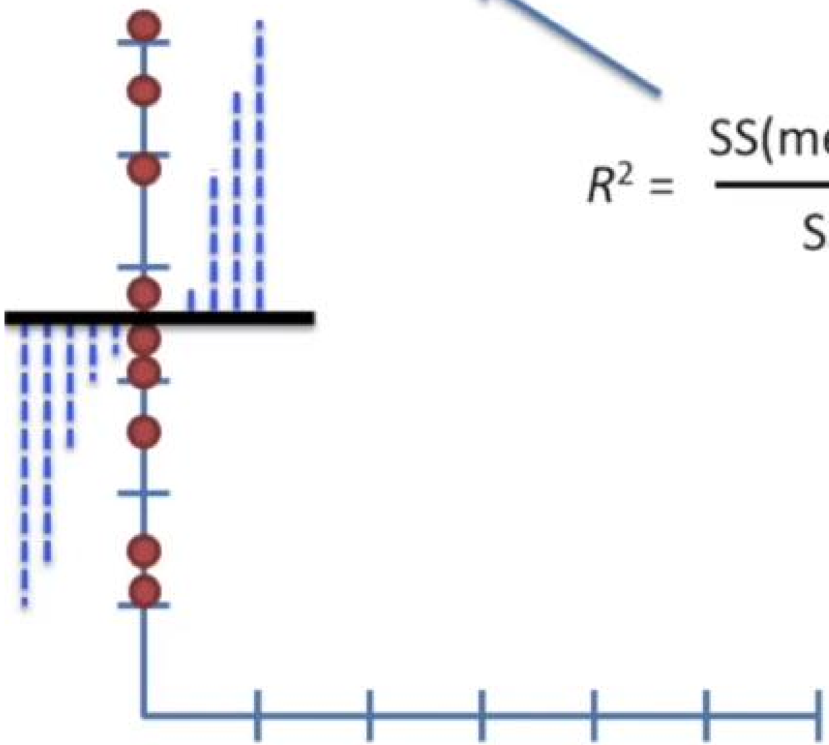


There is a 60% reduction in variance when we take the mouse weight into account.

Alternatively, we can say that mouse weight “explains” 60% of the variation in mouse size.

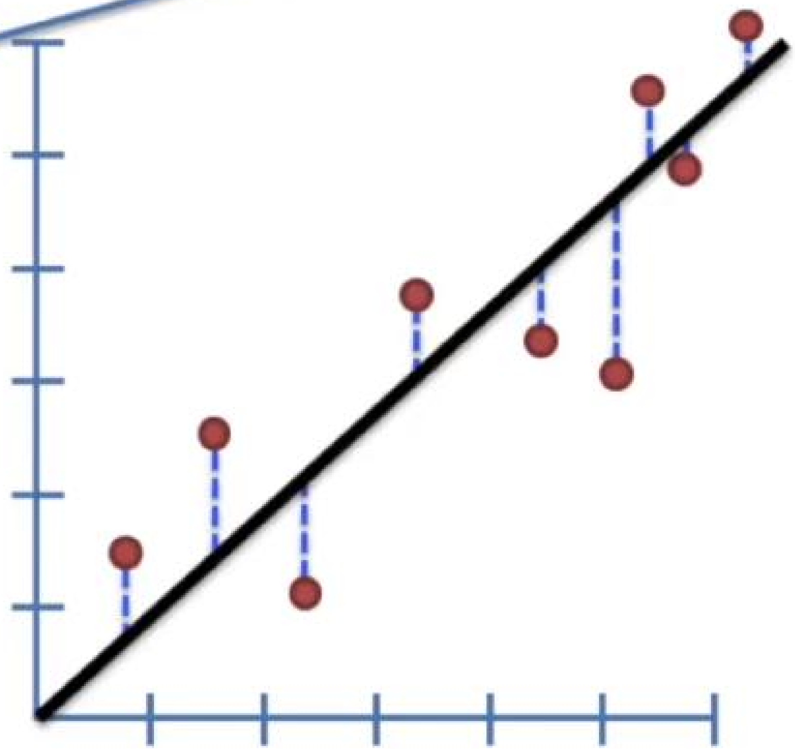
We can also use the sums of squares to make the same calculation.

SS(mean) = 100



$$R^2 = \frac{SS(\text{mean}) - SS(\text{fit})}{SS(\text{mean})}$$

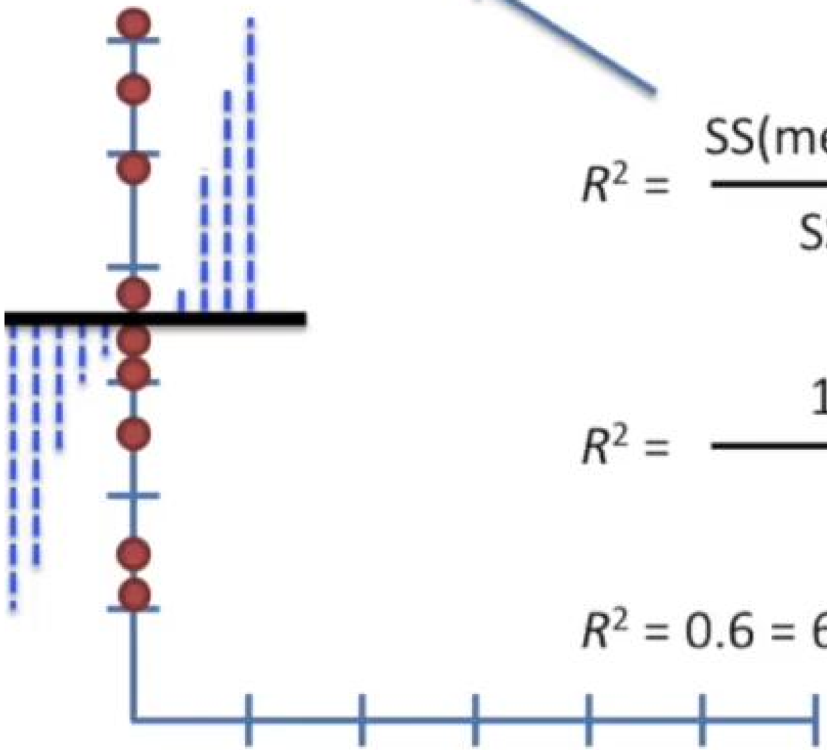
SS(fit) = 40



We can also use the sums of squares to make the same calculation.

SS(mean) = 100

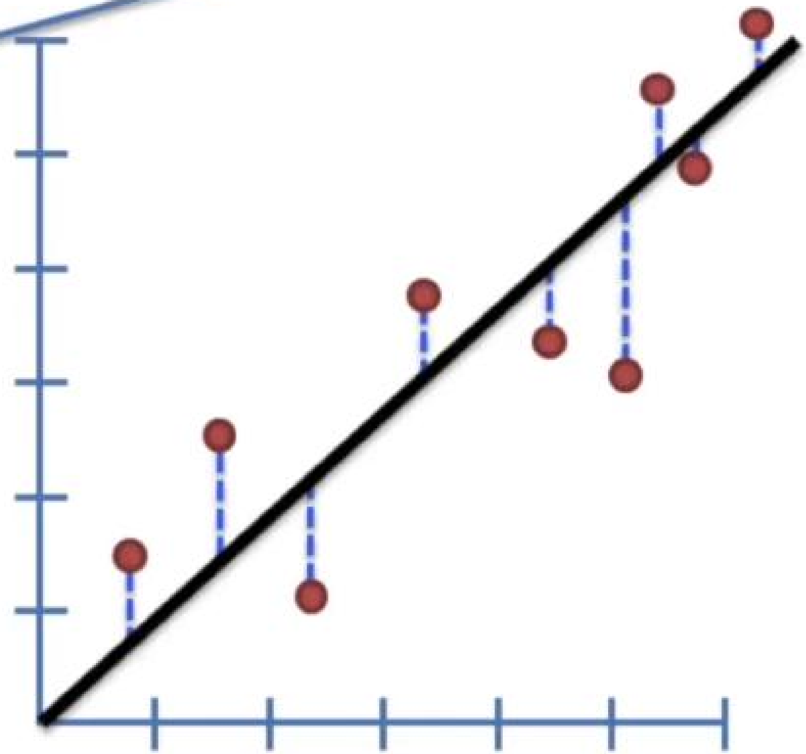
SS(fit) = 40



$$R^2 = \frac{SS(mean) - SS(fit)}{SS(mean)}$$

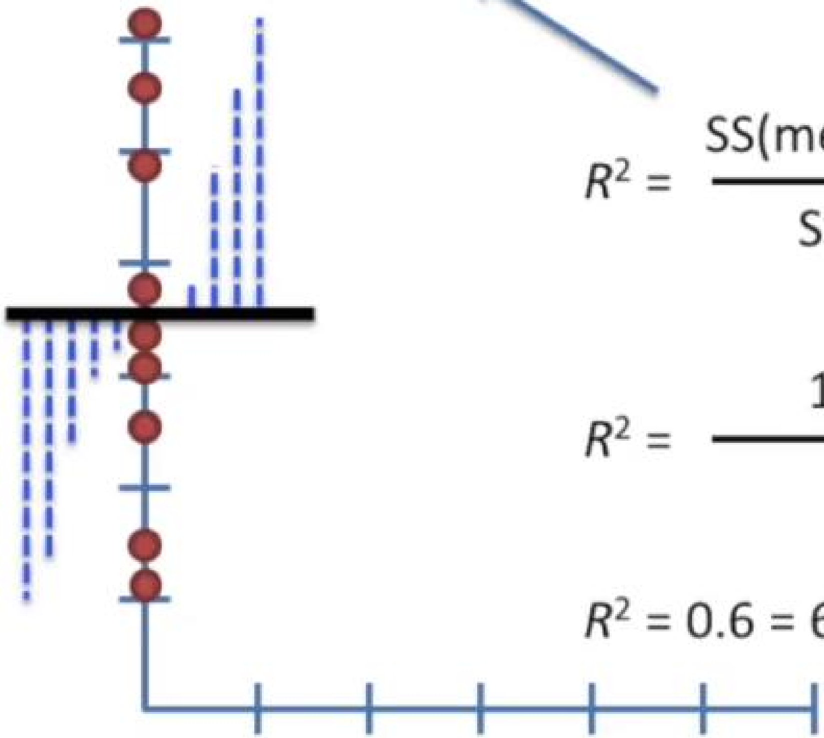
$$R^2 = \frac{100 - 40}{100}$$

$$R^2 = 0.6 = 60\%$$



We can also use the sums of squares to
make the same calculation.

SS(mean) = 100

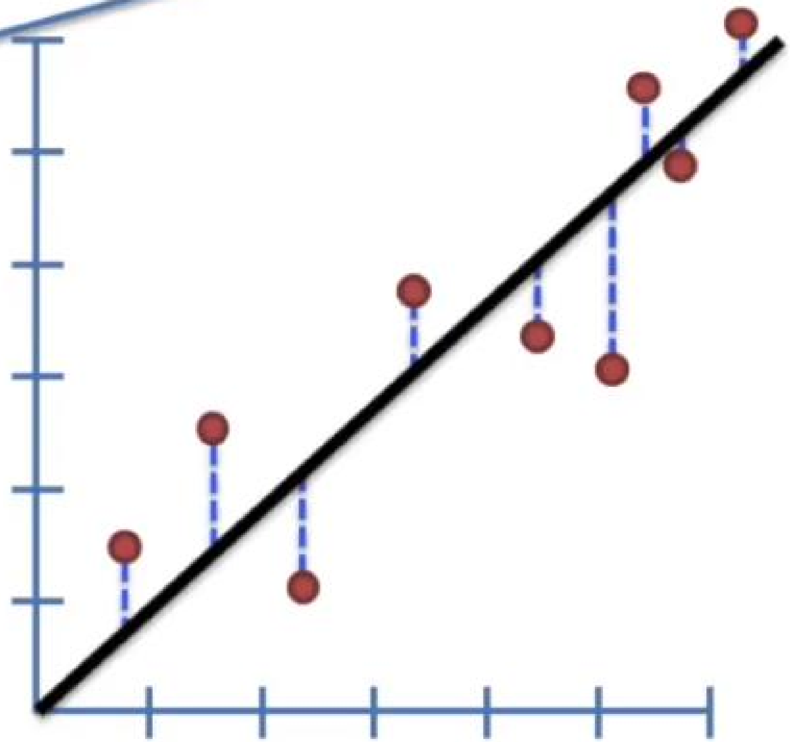


$$R^2 = \frac{SS(\text{mean}) - SS(\text{fit})}{SS(\text{mean})}$$

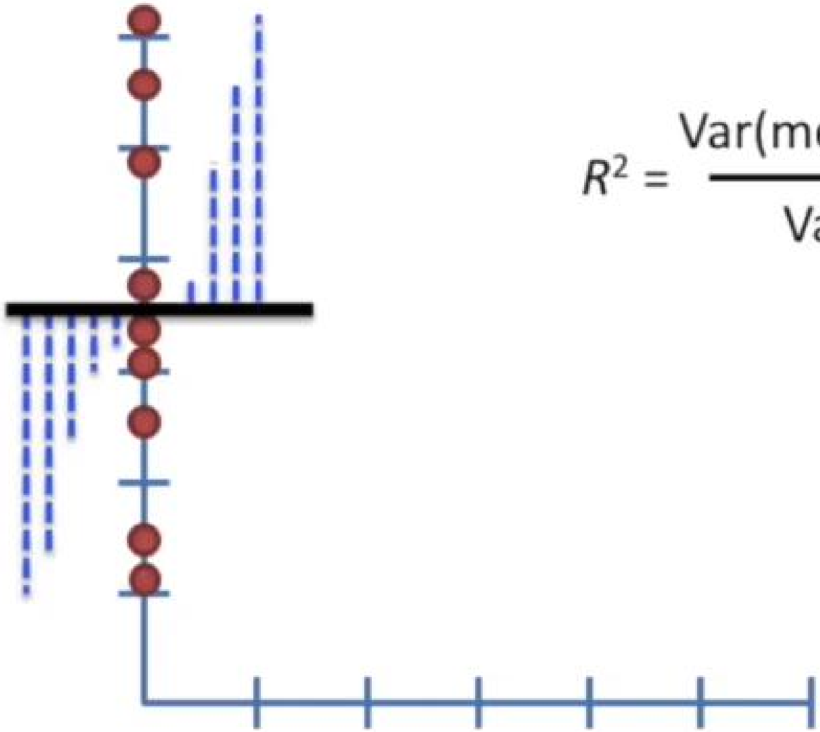
$$R^2 = \frac{100 - 40}{100}$$

$$R^2 = 0.6 = 60\%$$

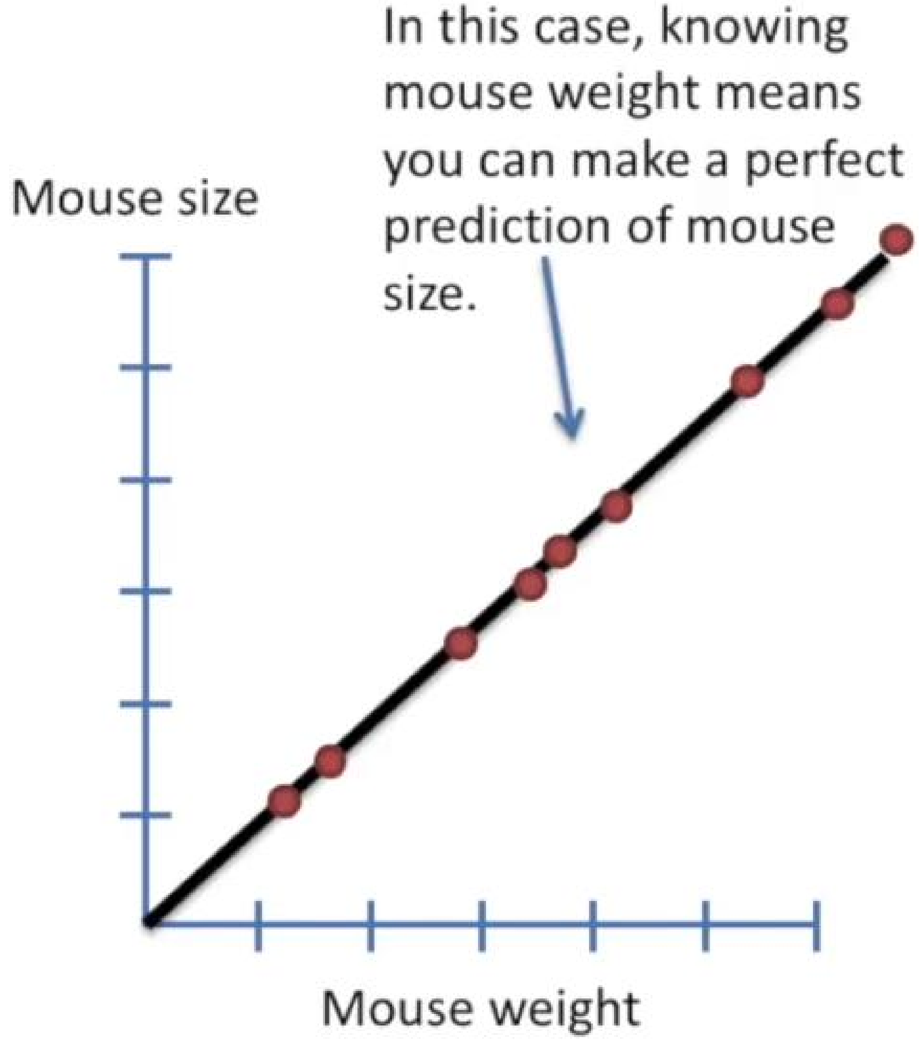
SS(fit) = 40



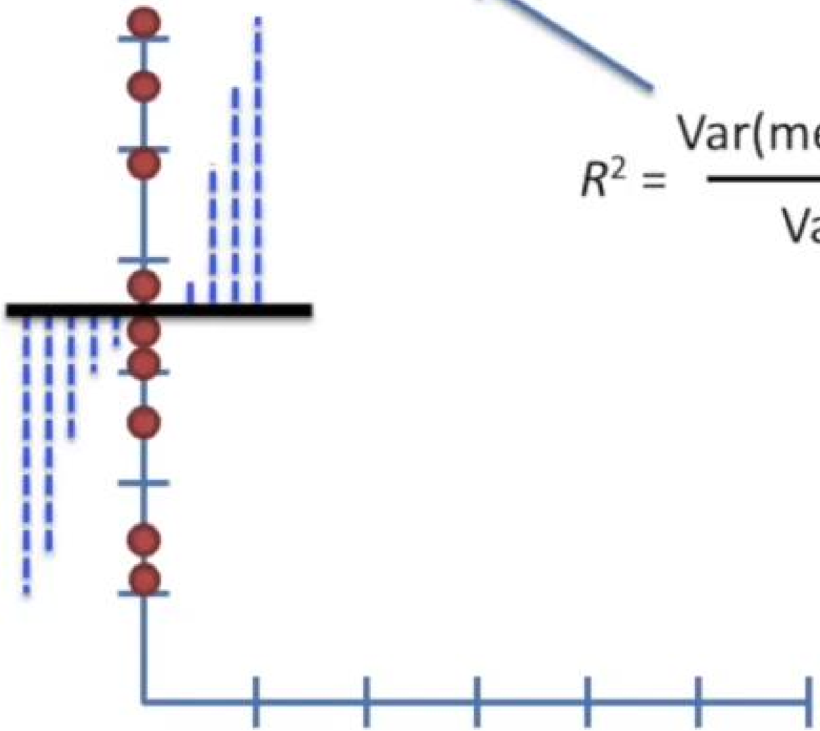
60% of the sums of squares of the
mouse size can be explained by
mouse weight..



$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

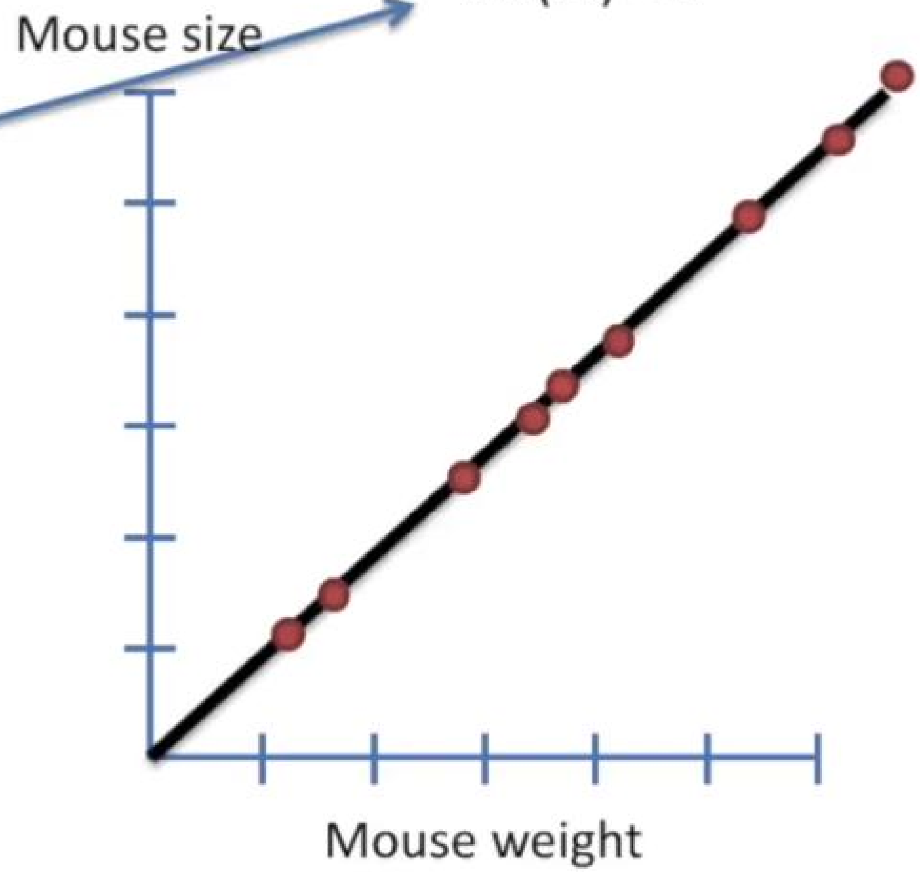


Var(mean) = 11.1



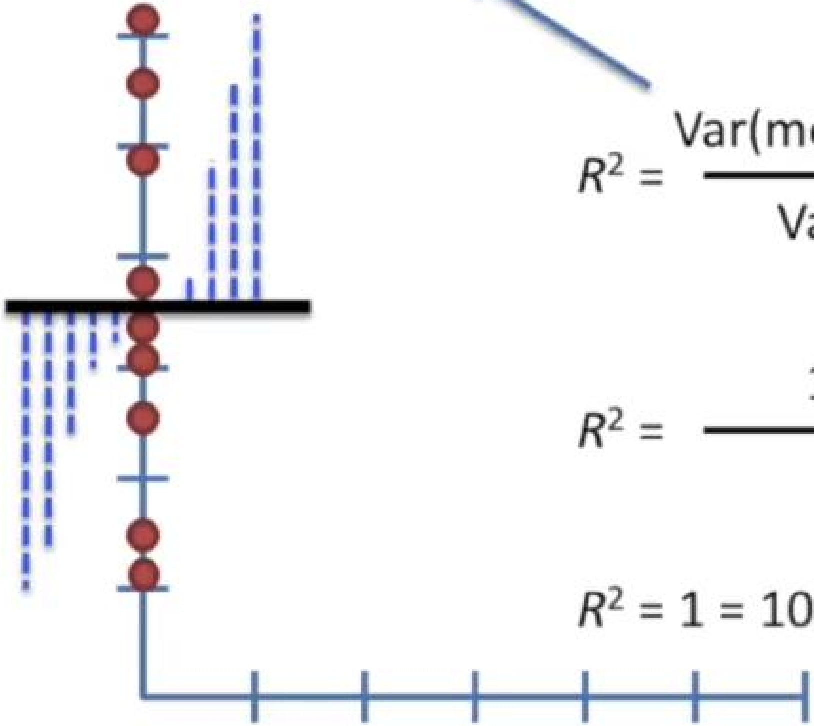
$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

Var(fit) = 0



Var(mean) = 11.1

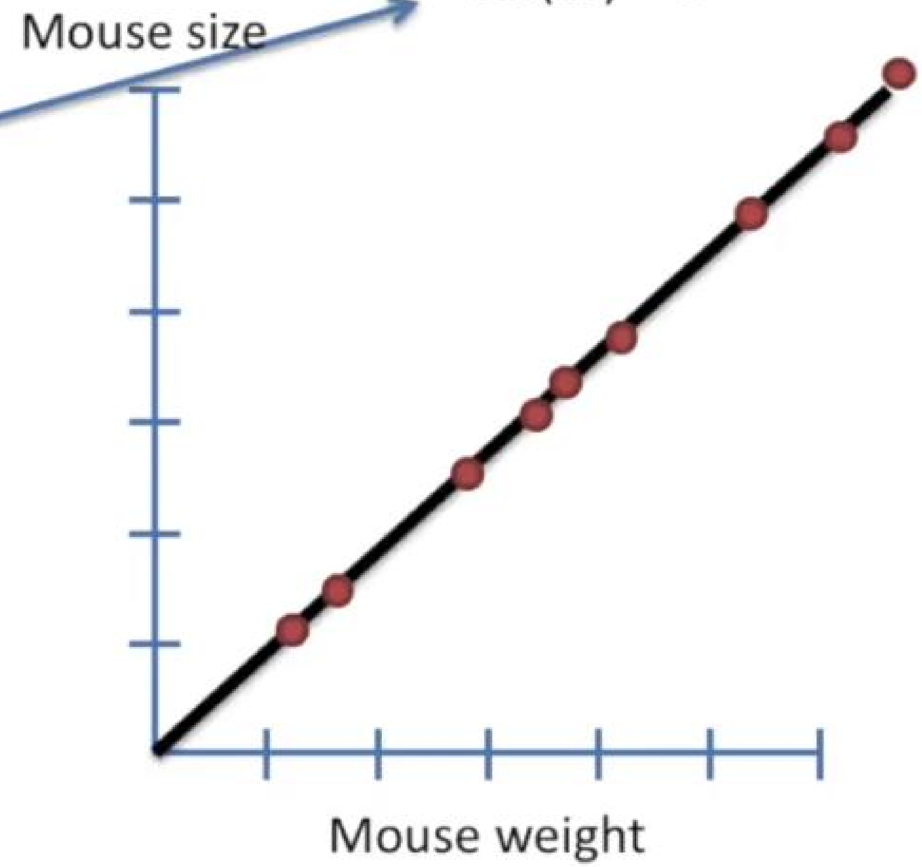
Var(fit) = 0



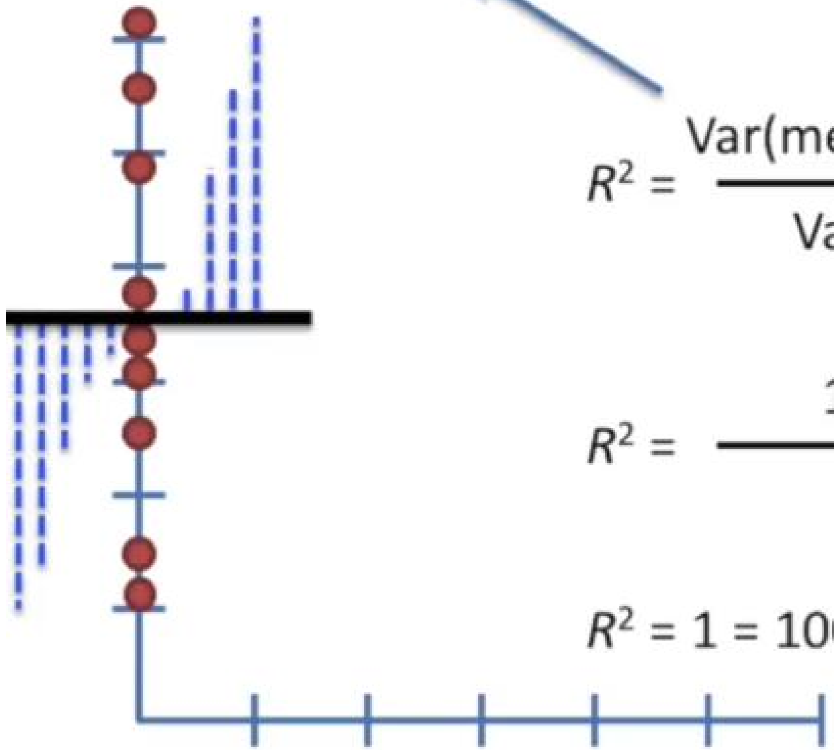
$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

$$R^2 = \frac{11.1 - 0}{11.1}$$

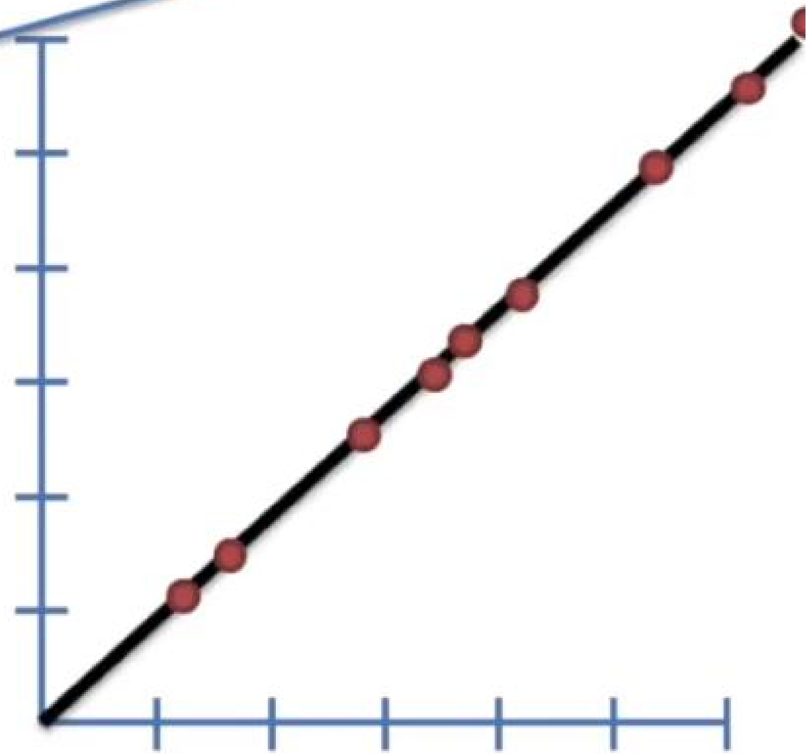
$$R^2 = 1 = 100\%$$



Var(mean) = 11.1



Var(fit) = 0



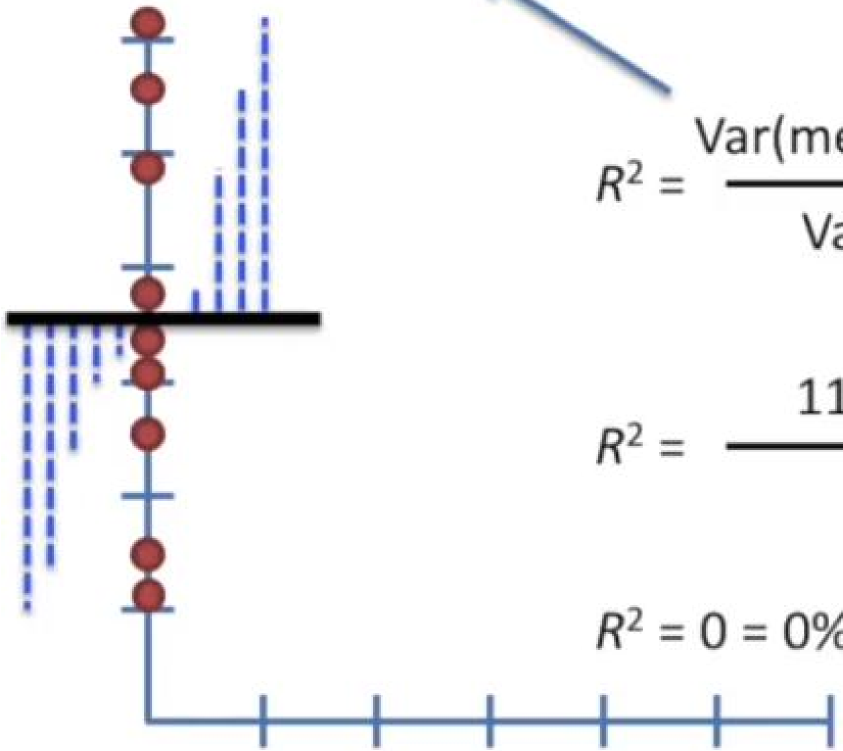
$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

$$R^2 = \frac{11.1 - 0}{11.1}$$

$$R^2 = 1 = 100\%$$

In this case, mouse weight "explains" 100% of the variation in mouse size.

Var(mean) = 11.1

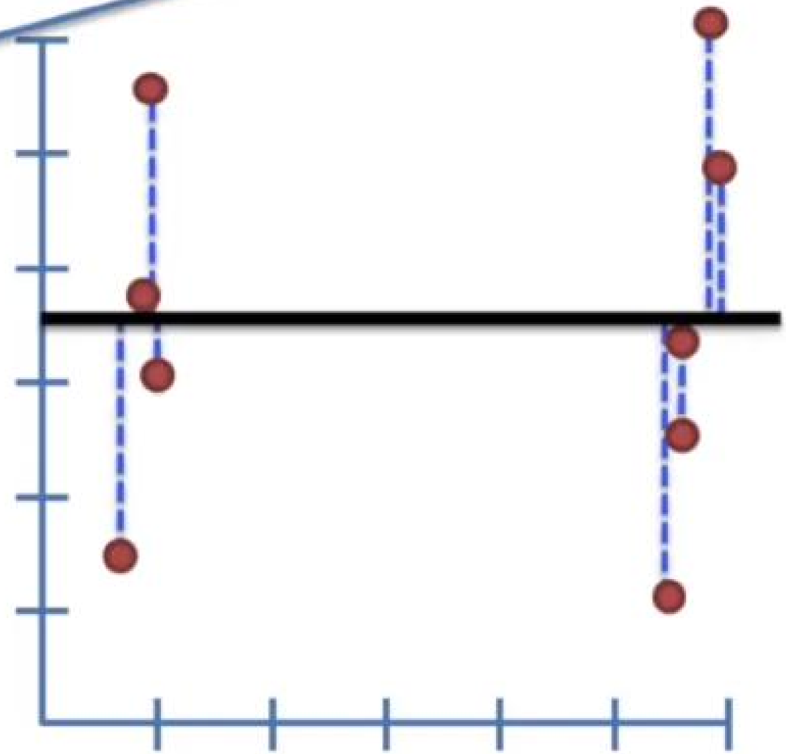


$$R^2 = \frac{\text{Var(mean)} - \text{Var(fit)}}{\text{Var(mean)}}$$

$$R^2 = \frac{11.1 - 11.1}{11.1}$$

$$R^2 = 0 = 0\%$$

Var(fit) = 11.1

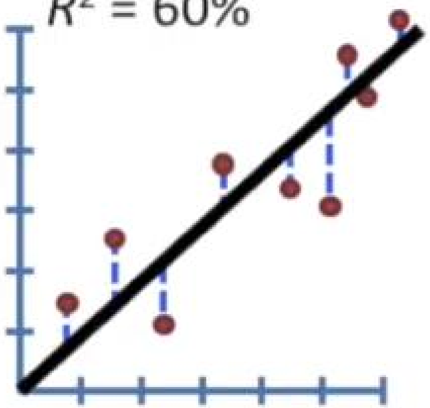


In this case, mouse weight doesn't "explain" any of the variation around the mean.

But the concept applies to any equation, no matter how complicated.

$y = 0.1 + 0.78x$

$R^2 = 60\%$



$y = 0.1 + 0.78x - 8.3z + ...$

1) Measure, square and sum the distance from the data to the mean.

2) Measure, square and sum the distance from the data to the complicated equation.

$$R^2 = \frac{SS(\text{mean}) - SS(\text{fit})}{SS(\text{mean})}$$

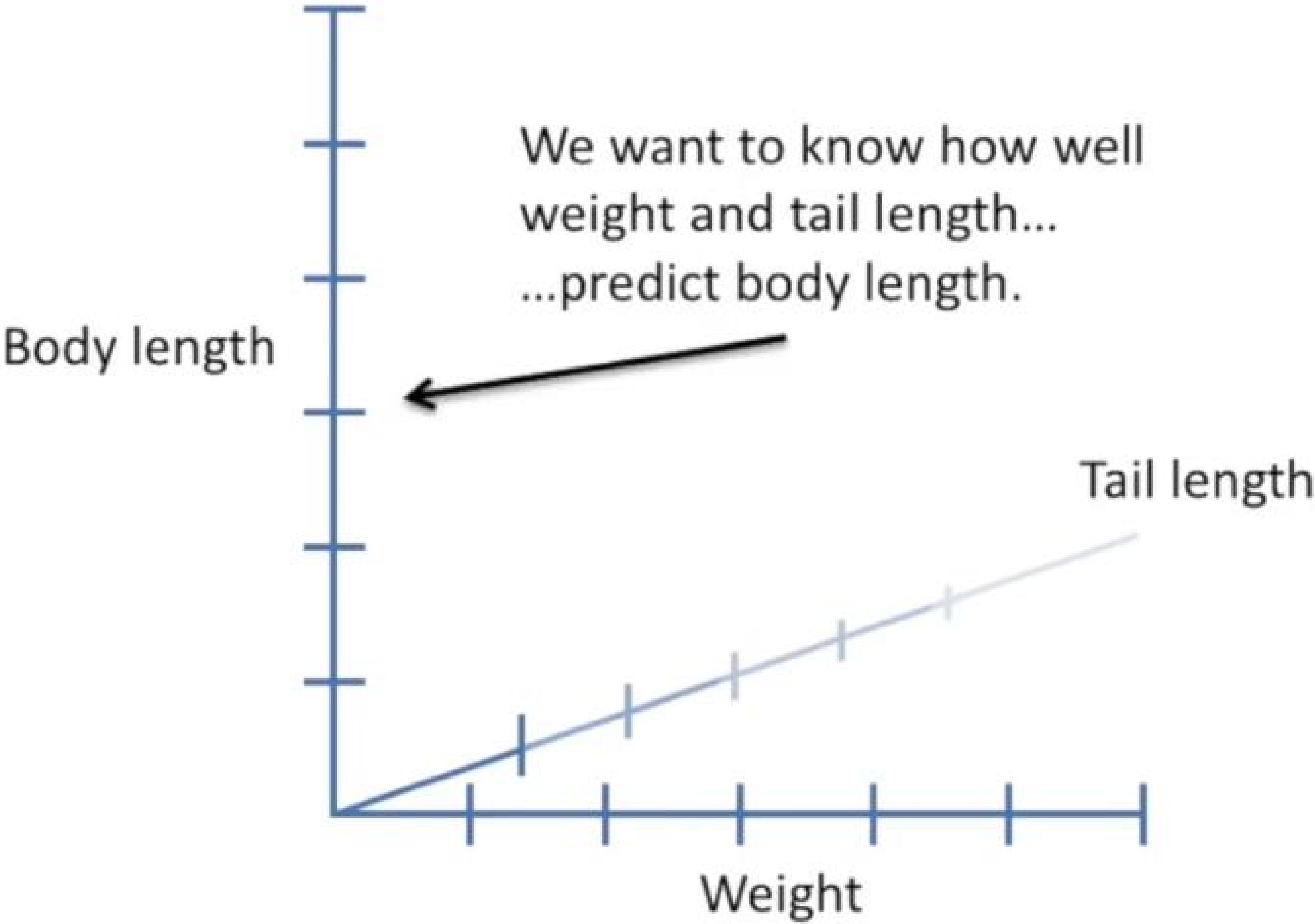
Let's look at slightly more complicated example:

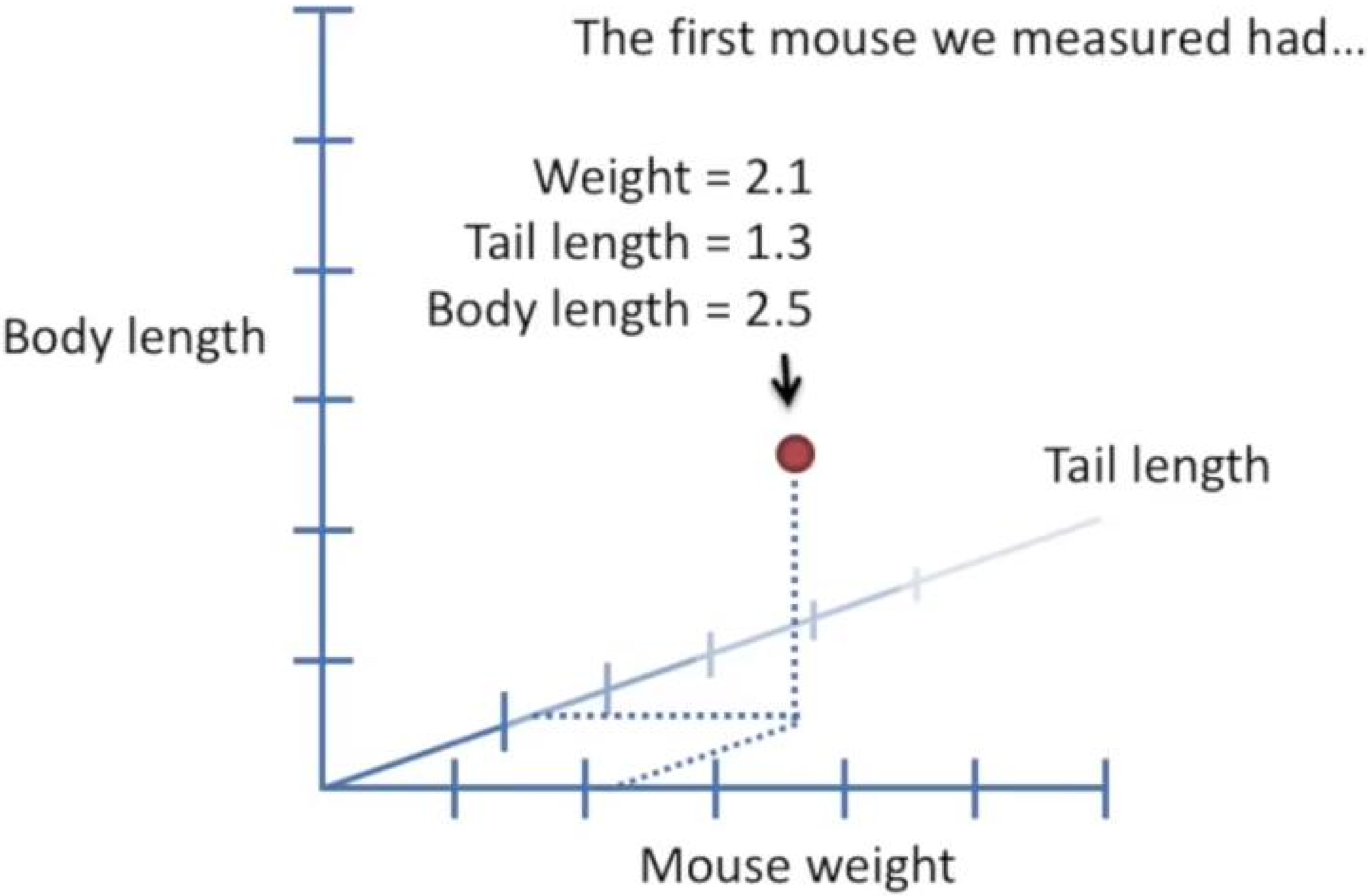
Imagine we wanted to know if mouse **weight** and **tail length** did a good job predicting the **length of the mouse's body**.

So we measured a bunch of mice…

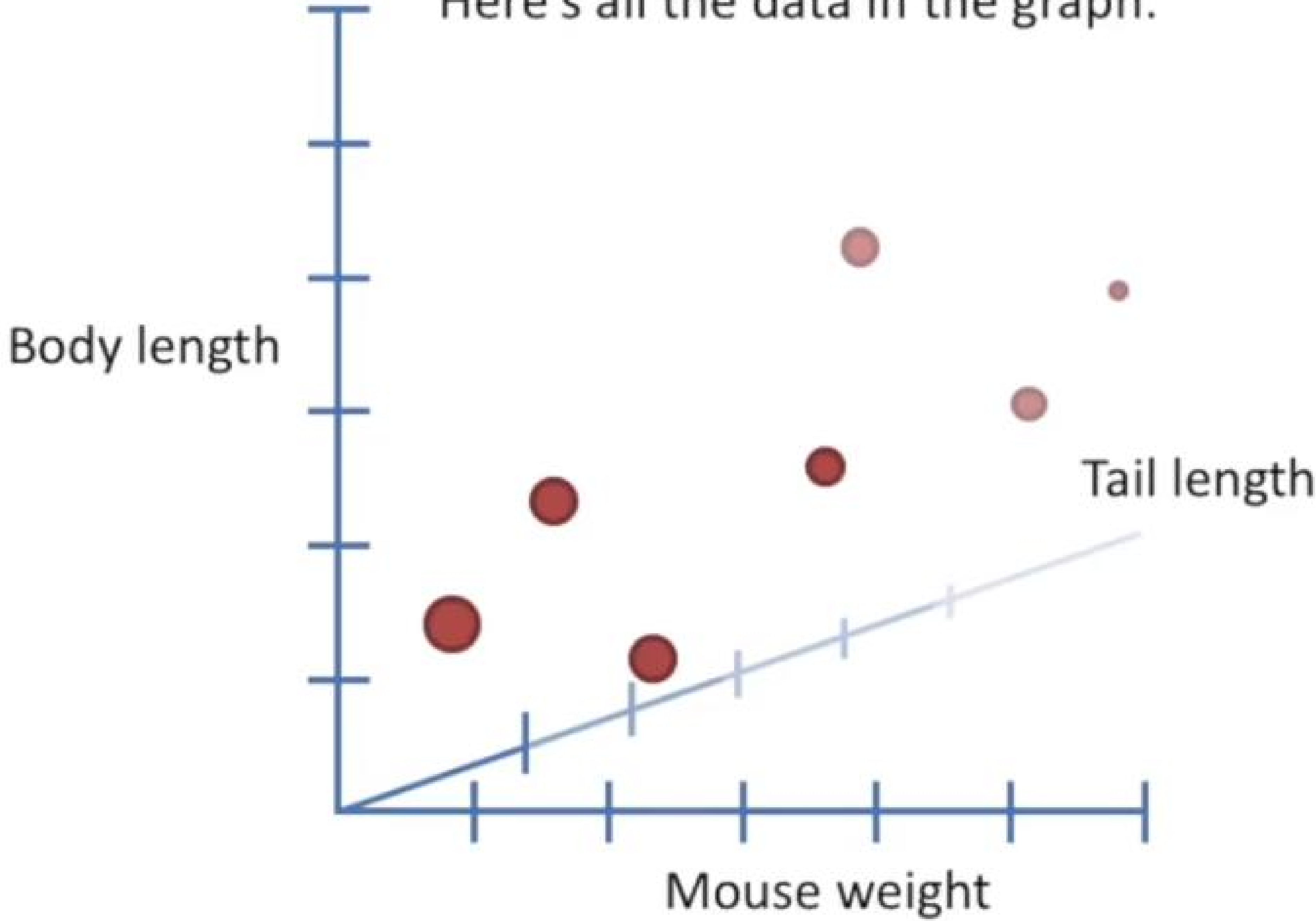
Weight	Tail Length	Body Length
3.5	2.9	3.1
1.3	2.1	2.8
5.9	4.1	6.1
4.8	3.2	3.8
…	…	…

To plot this data, we need a 3-dimensional graph.

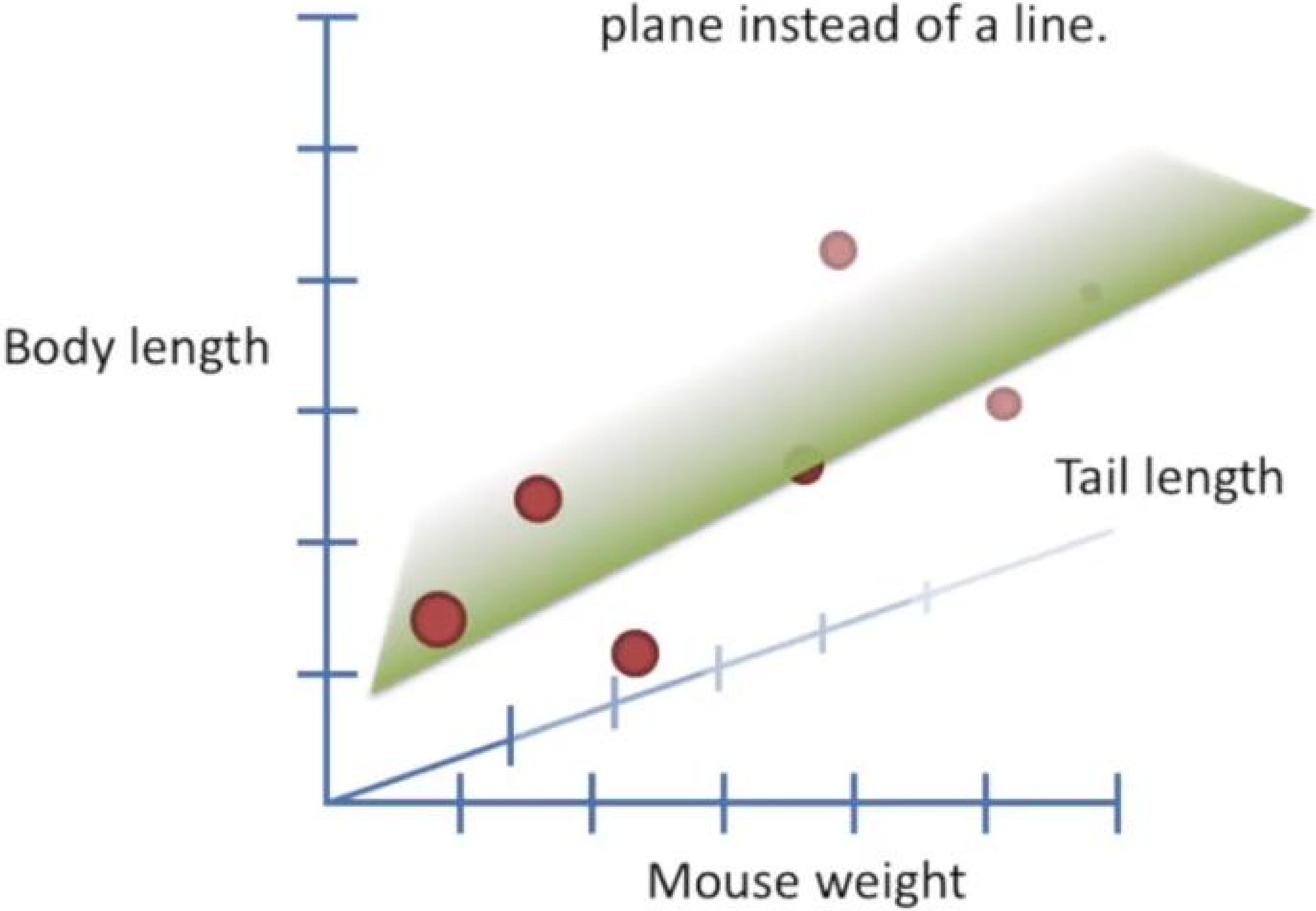



The first mouse we measured had...
Weight = 2.1
Tail length = 1.3
Body length = 2.5
Body length
Tail length
Mouse weight

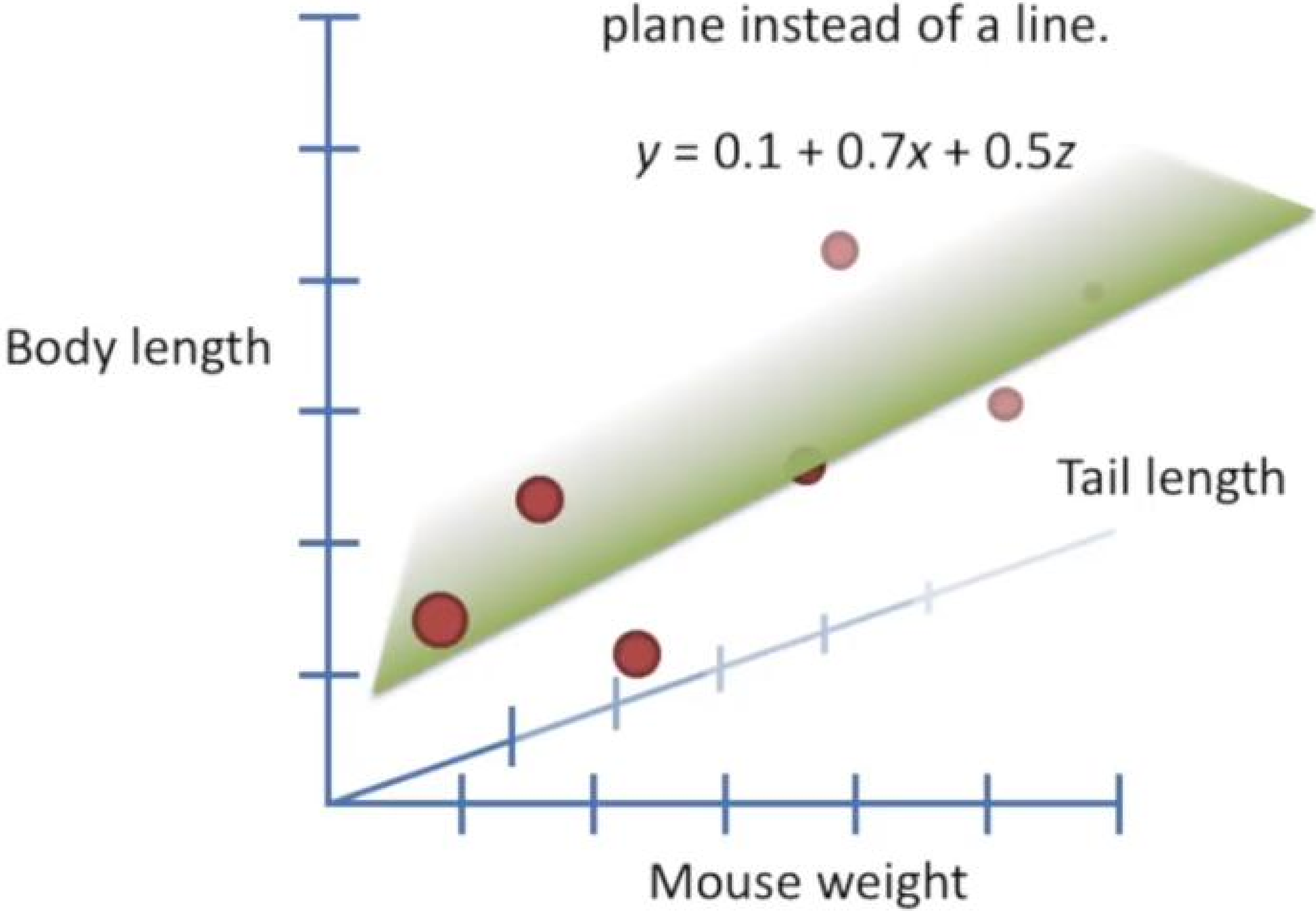
Here's all the data in the graph.



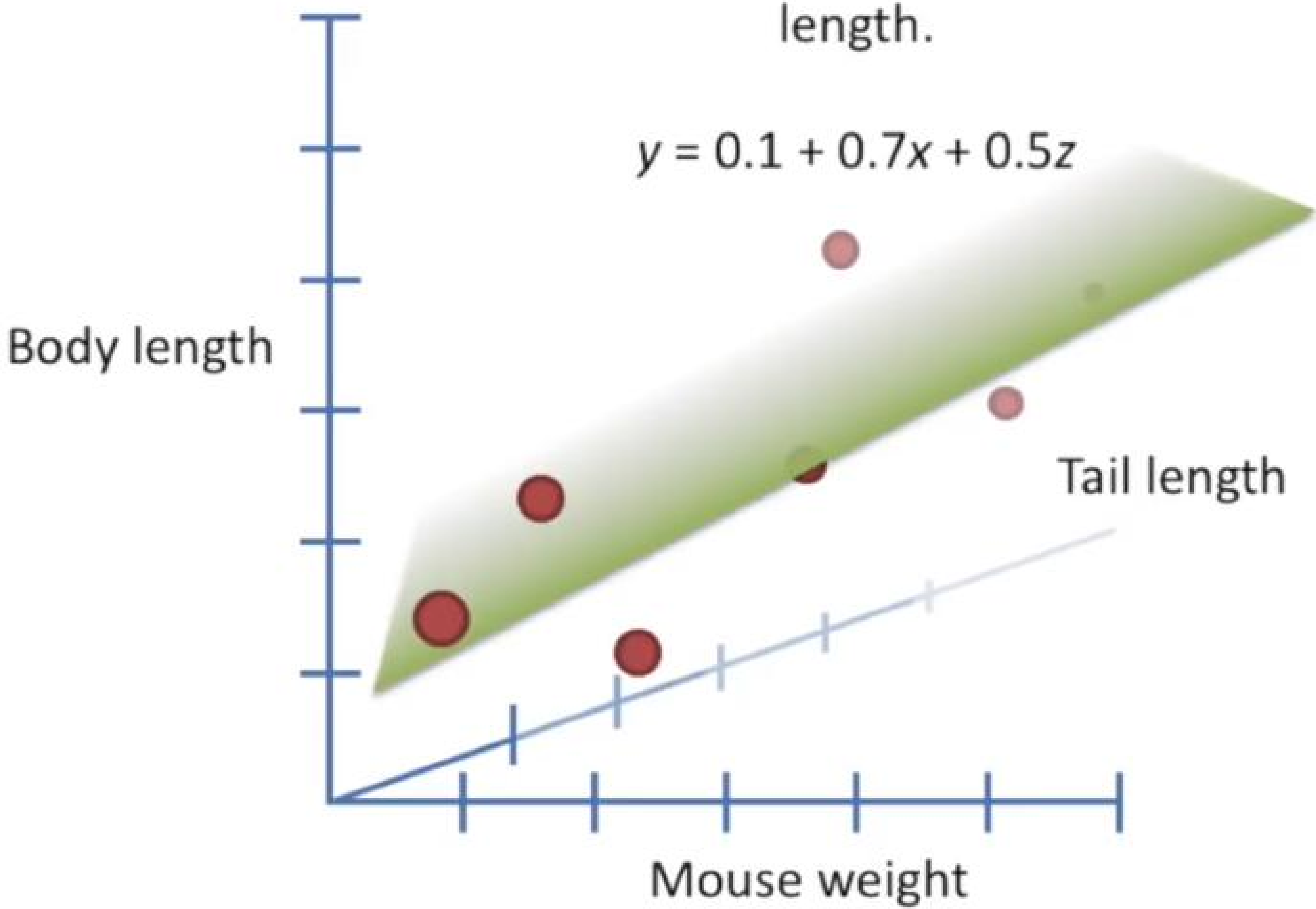
Now we do a least-squares fit. Since we have the extra term in the equation, we fit a plane instead of a line.

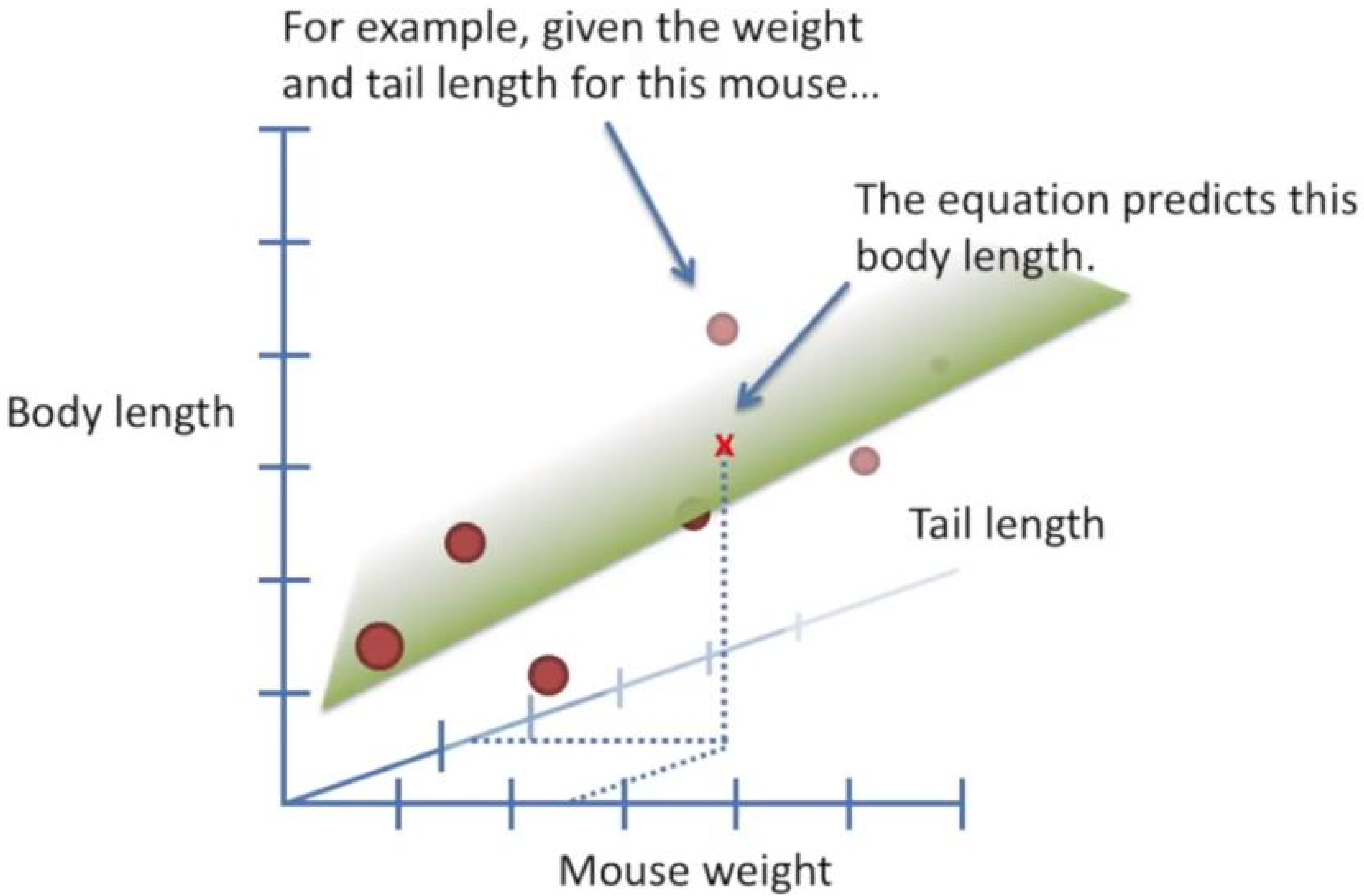


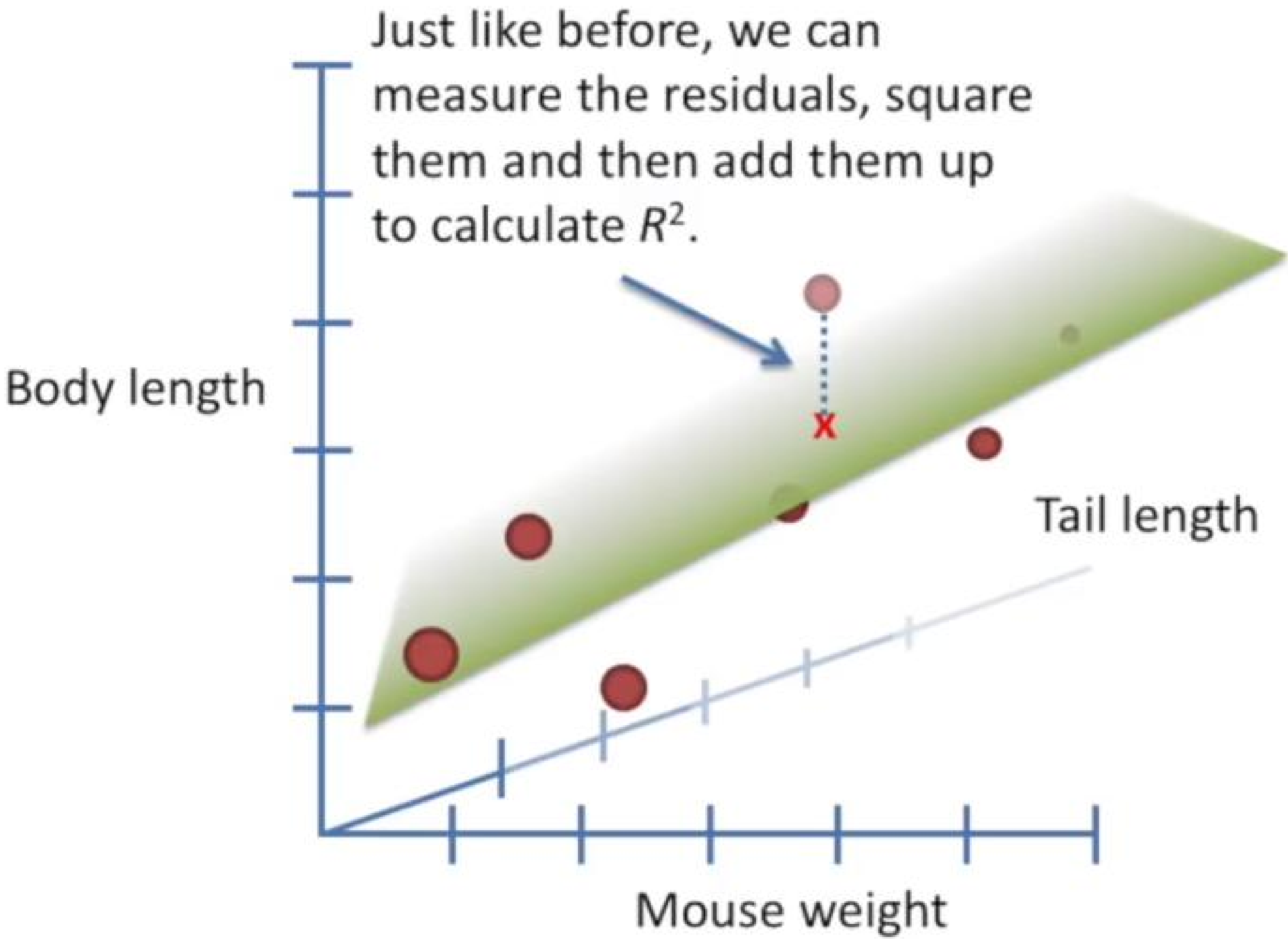
Now we do a least-squares fit. Since we have the extra term in the equation, we fit a plane instead of a line.



If we know a mouse's weight and tail length, we can use the equation to guess the body length.

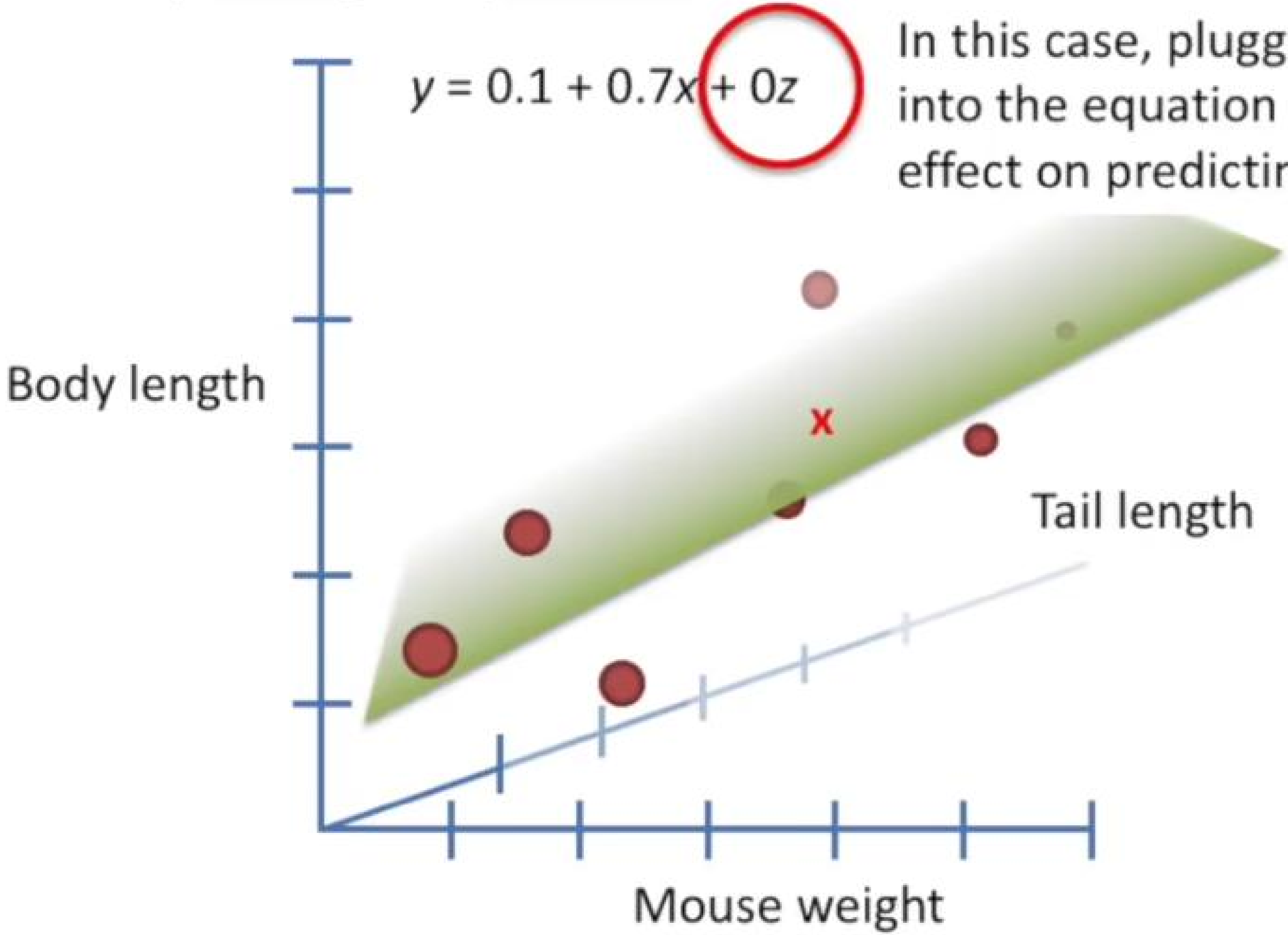



For example, given the weight and tail length for this mouse...
The equation predicts this body length.
Body length
Tail length
Mouse weight

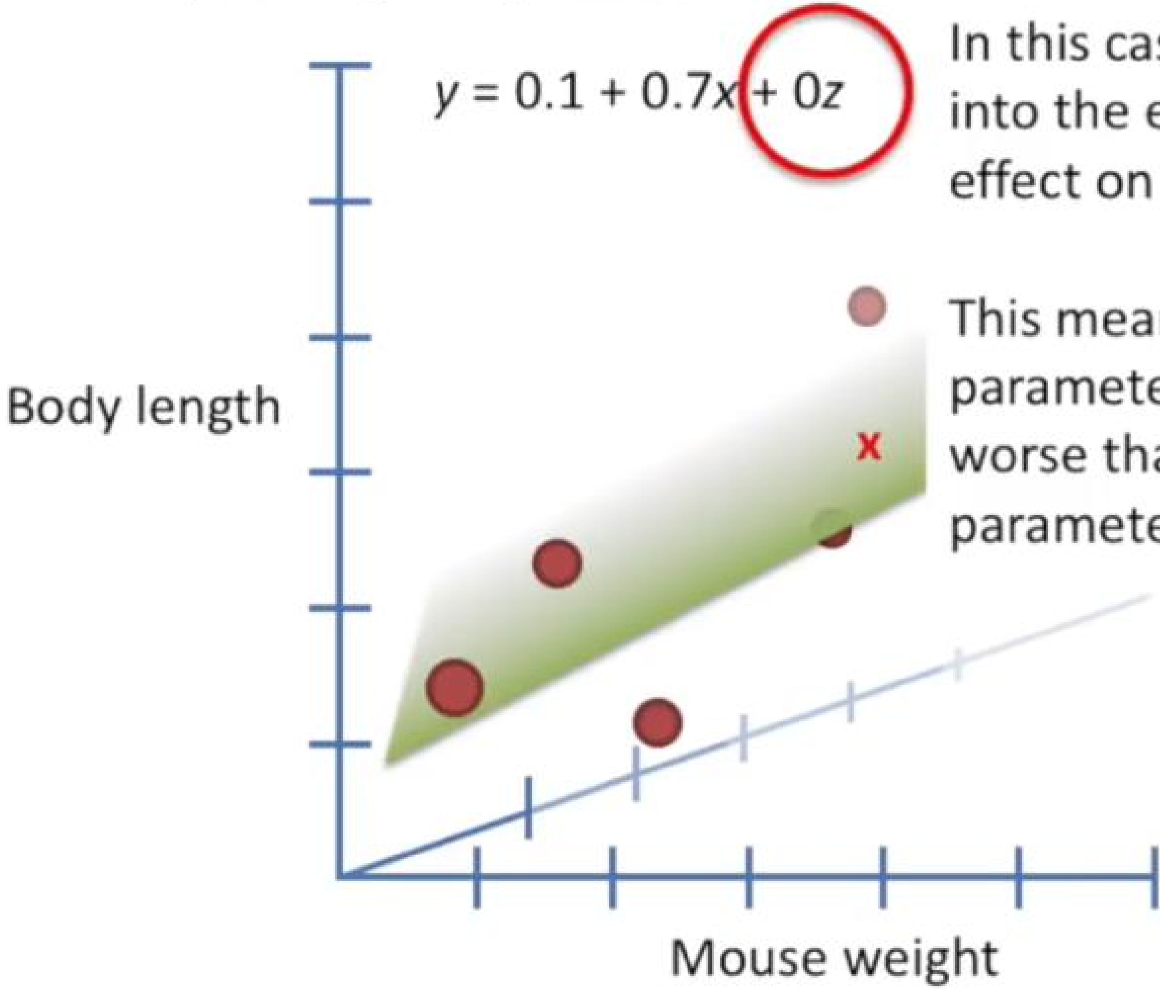

Just like before, we can measure the residuals, square them and then add them up to calculate R^2.
Body length
Tail length
Mouse weight

Now, if the tail length (z-axis) is useless and doesn't make **SS(fit)** smaller, then least-squares will ignore it by making that parameter = 0.

In this case, plugging the tail length into the equation would have no effect on predicting the mouse size.



Now, if the tail length (z-axis) is useless and doesn't make **SS(fit)** smaller, then least-squares will ignore it by making that parameter = 0.



In this case, plugging the tail length into the equation would have no effect on predicting the mouse size.

This means equations with more parameters will never make **SS(fit)** worse than equations with fewer parameters.

In other words:

This equation...

Mouse size = 0.3 + mouse weight + flip of a coin + favorite color + astrological sign +....

... will never perform worse than this equation...

Mouse size = 0.3 + mouse weight

In other words:

This equation...

Mouse size = 0.3 + mouse weight + 0×flip of a coin + 0×favorite color + ...

... will never perform worse than this equation...

Mouse size = 0.3 + mouse weight

This is because least squares will cause any term that makes SS(fit) worse to be multiplied by 0, and, in a sense, no longer exist.

This equation…

Mouse size = 0.3 + mouse weight + flip of a coin

Now, due to random chance, there is a small probability that the small mice in the dataset might get heads more frequently than large mice.

If this happened, then we'd get a smaller **SS(fit)**, and a better R^2

This equation...

Mouse size = 0.3 + mouse weight + flip of a coin + favorite color + astrological sign +...

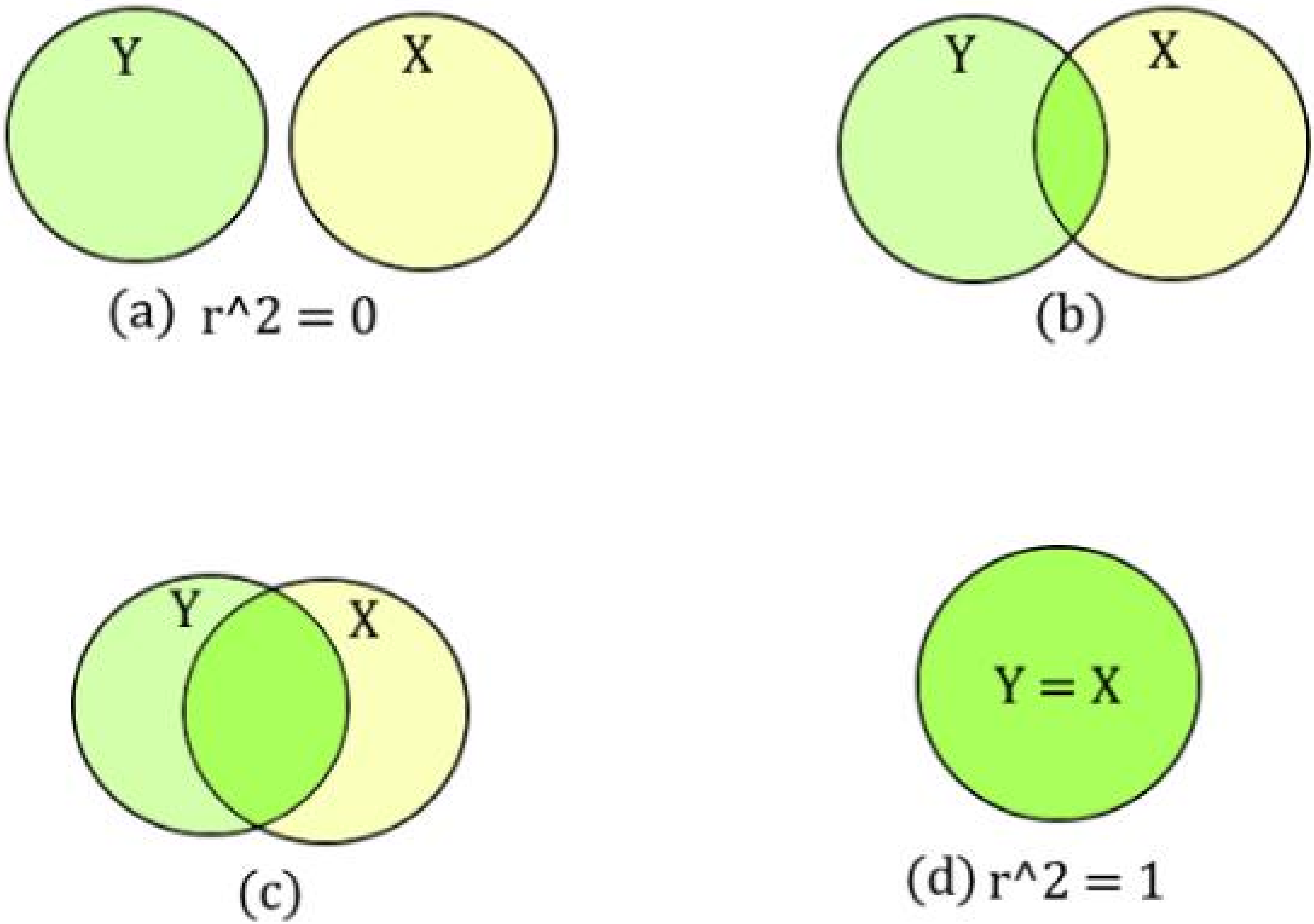
The more parameters we add to the equation, the more opportunities we have for random events to reduce **SS(fit)** and result in a better R^2

This equation...

Mouse size = 0.3 + mouse weight + flip of a coin + favorite color + astrological sign +...

The more parameters we add to the equation, the more opportunities we have for random events to reduce **SS(fit)** and result in a better R^2

Thus, people report an "adjusted R^2" value that, in essence, scales R^2 by the number of parameters.


Y
X
(a) r^2 = 0
Y
X
(b)
Y
X
(c)
Y = X
(d) r^2 = 1

The overlap of the circles represents the **extent to which the variation in Y is explained by the variation in X.**

Problem with R^2 — Value increases with the number of explanatory variables

Think about it. R^2 is the ratio of the explained variance to the total variance. On adding a new variable the explained variance and hence the value of R^2 ***will increase, or at least, will not decrease***.

However, this ***does not at all*** mean that the model with the added variable is better than the model without it. R^2 can be misleading if used to compare models with a different number of predictors.

Adjusted R^2 is a modified version of R^2 adjusted with the number of predictors. It penalizes for adding unnecessary features and allows a comparison of regression models with a different number of predictors.

$$Adjusted\ R^2 = \bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

Here $\boldsymbol{k}$ is the number of explanatory variables in the model and $\boldsymbol{n}$ is the number of observations.

The value of adjusted R^2 is always less than that of R^2.

The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance.

Also, note that the value of adjusted R^2 can be negative.

Obtaining a negative value for Adjusted R^2 can indicate few or all of the following:

- The linear model is a poor fit for the data
- The number of predictors is large
- The number of samples is small