Probabilistic modelling

Probabilistic modelling is a modelling approach that allows us to take uncertainty into account. It is based on the theory of probability, which considers randomness in predicting future events.

A probabilistic model gives a distribution of possible outcomes. Meaning it defines all outcomes and allows for a certain degree of likelihood for each outcome to occur. Probabilistic modelling and reasoning is particularly valuable when studying population health, as it can be used to predict and improve health outcomes of a population.

# factors that could influence the health of a population

* Environmental; such as living and working conditions,
* Social and economic; such as, education, income and employment,
* Lifestyle; what people eat and drink, whether they smoke, and how much physical activity they do
* Access to healthcare.

# Correlation

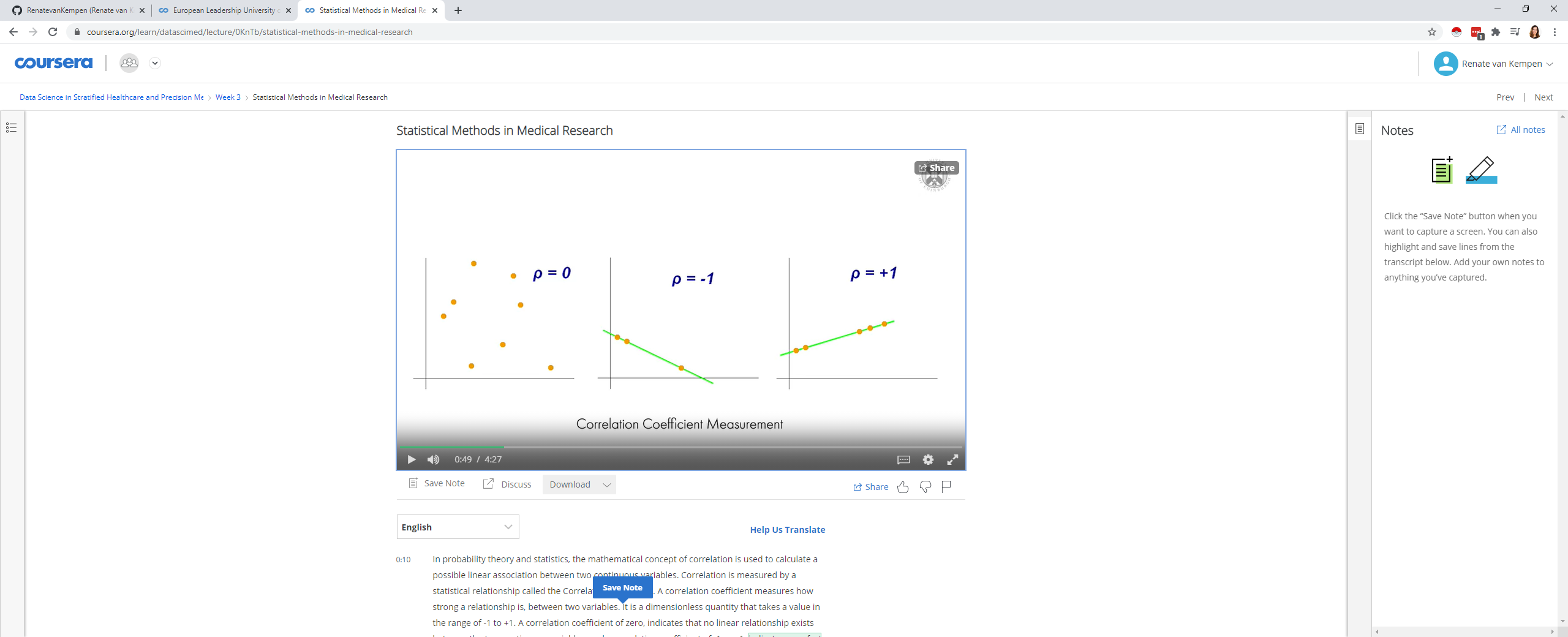
the mathematical concept of correlation is used to calculate

a possible linear association between two continuous variables.

Correlation is measured by a statistical relationship called the Correlation Coefficient.

## Correlation coefficient

A correlation coefficient measures how strong a relationship is, between two variables



It is a dimensionless quantity that takes a value in the range of -1 to +1.

A correlation coefficient of zero,

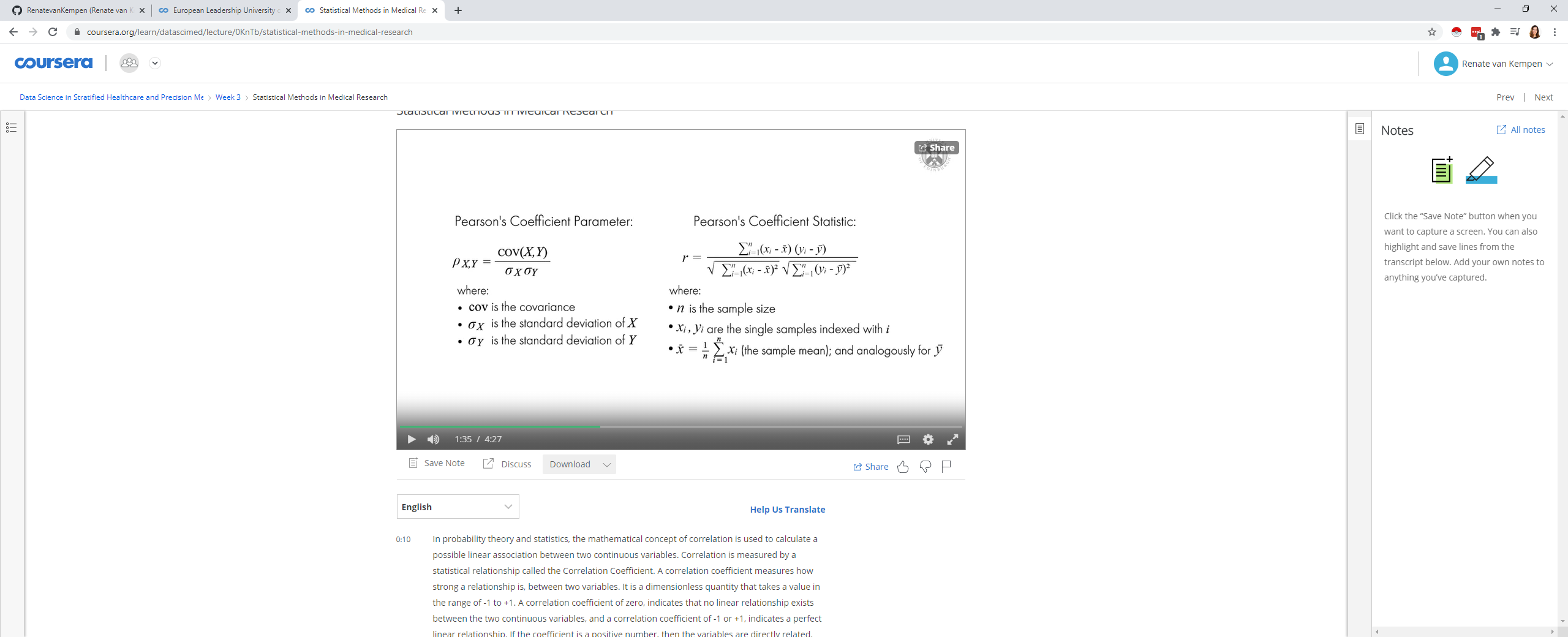
indicates that no linear relationship exists between the two continuous variables,

and a correlation coefficient of -1 or +1,

indicates a perfect linear relationship.

If the relationship between two continuous variables is not linear, it is not correlation in statistical terms.

## Pearsons coefficient parameter



A well known and commonly used type of correlation coefficient is Pearson's product-moment correlation coefficient. It is denoted as rho for a population parameter, and as r for a sample statistic. The difference between a statistic and a parameter, is that a statistic describes a sample, whereas, a parameter describes an entire population. This coefficient is a measure of the strength and direction of the linear relationship between two variables, that is defined as the covariance of the variables divided by the product of their standard deviations. It is used when both variables being studied are normally distributed.

# REgression analysis

Regression analysis uses mathematical models to describe relationships.

The main difference between correlation analysis and regression analysis,

is that correlation analysis focuses primarily on association,

while regression analysis is used to make predictions.

Simple linear regression is used to calculate the relationship between two variables.

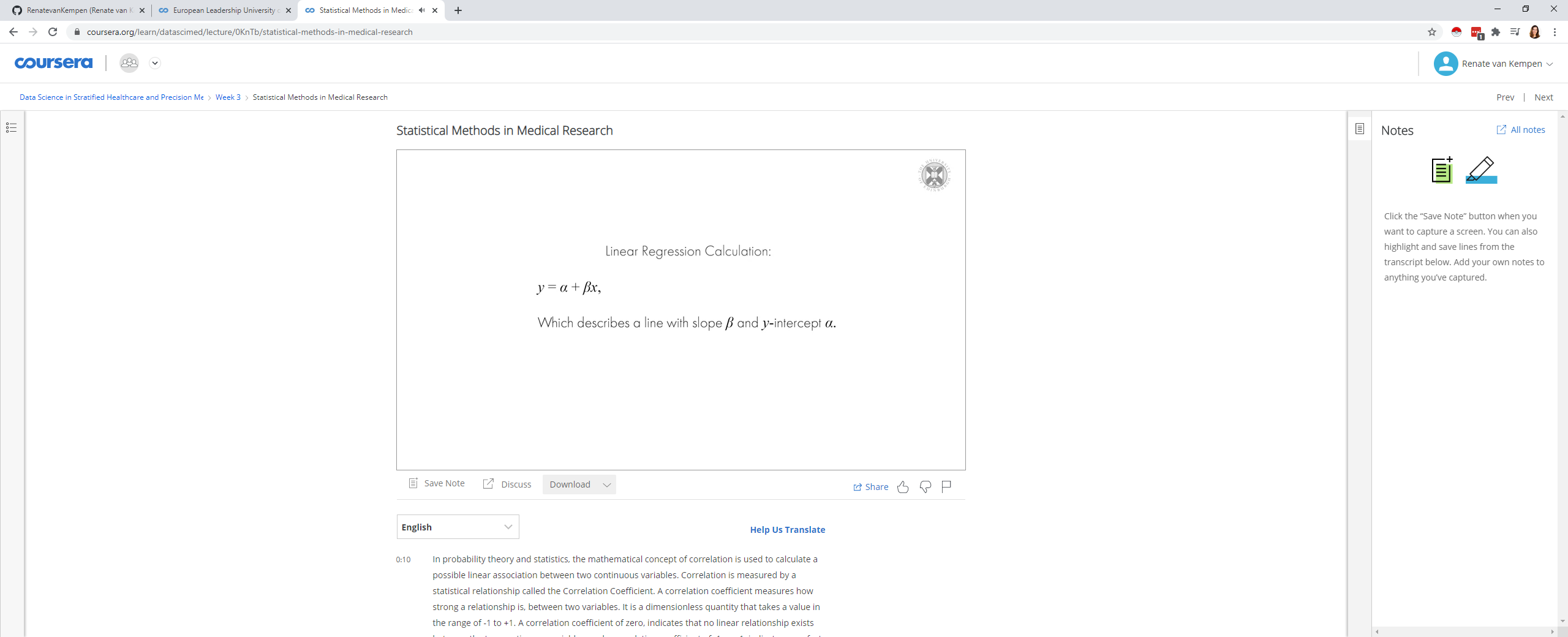
Where one variable, the dependent variable, denoted by y,

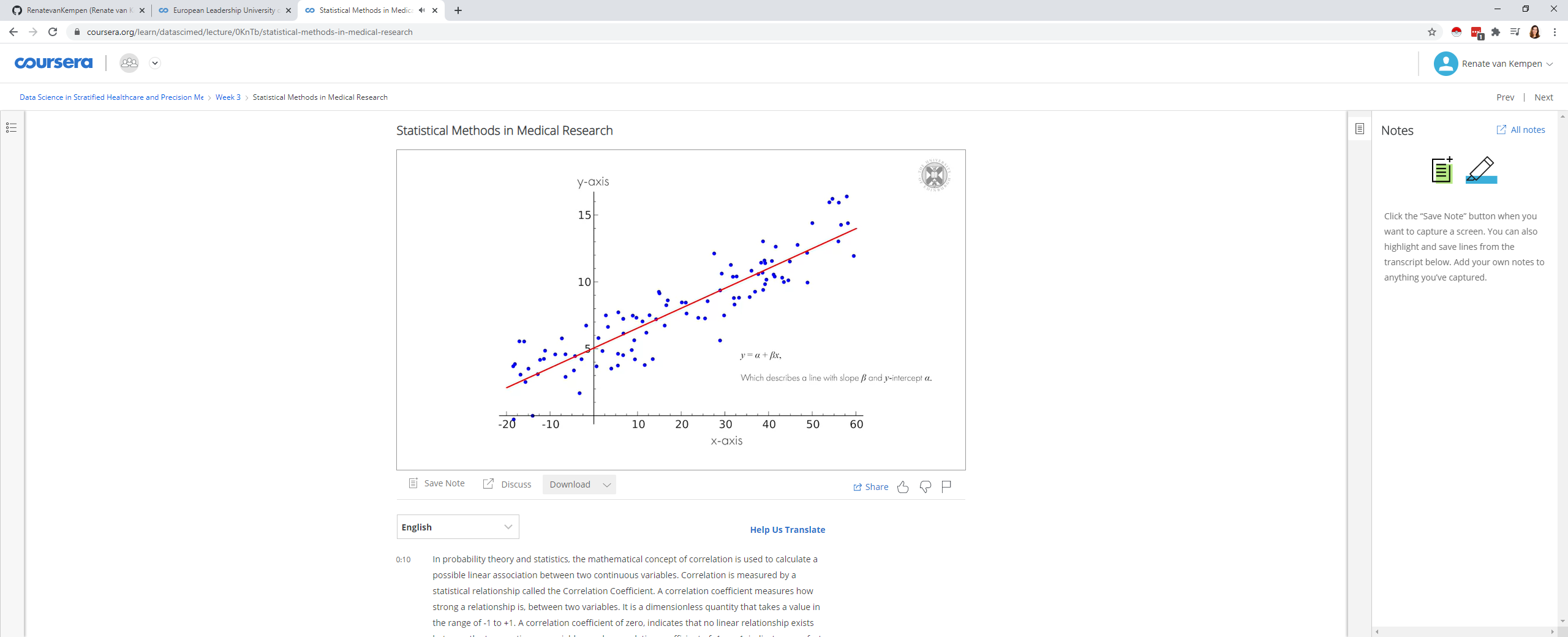
is expected to change as the other variable,

the independent variable, denoted by x, changes.

This technique fits a straight line to data,

where it is the so-called regression line.





## scatter plots

Scatter plots are similar to line graphs,

in that they use horizontal and vertical axes to plot data points.

Scatter plots are important in statistics,

because they can show the extent of correlation if any, between two continuous variables.

