Start-Tech Academy

Correlation

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

Correlation

Examples

Some examples of data that have a high correlation:
- Your caloric intake and your weight.
- The amount of time your study and your GPA.

Some examples of data that have a low correlation (or none at all):
- A dog's name and the type of dog biscuit they prefer.
- The cost of a car wash and how long it takes to buy a soda inside the station.

The Correlation Coefficient

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

Correlation Coefficient

Definition

- A correlation coefficient is a way to put a value to the relationship.
- Correlation coefficients have a value of between -1 and 1.
- A "0" means there is no relationship between the variables at all,
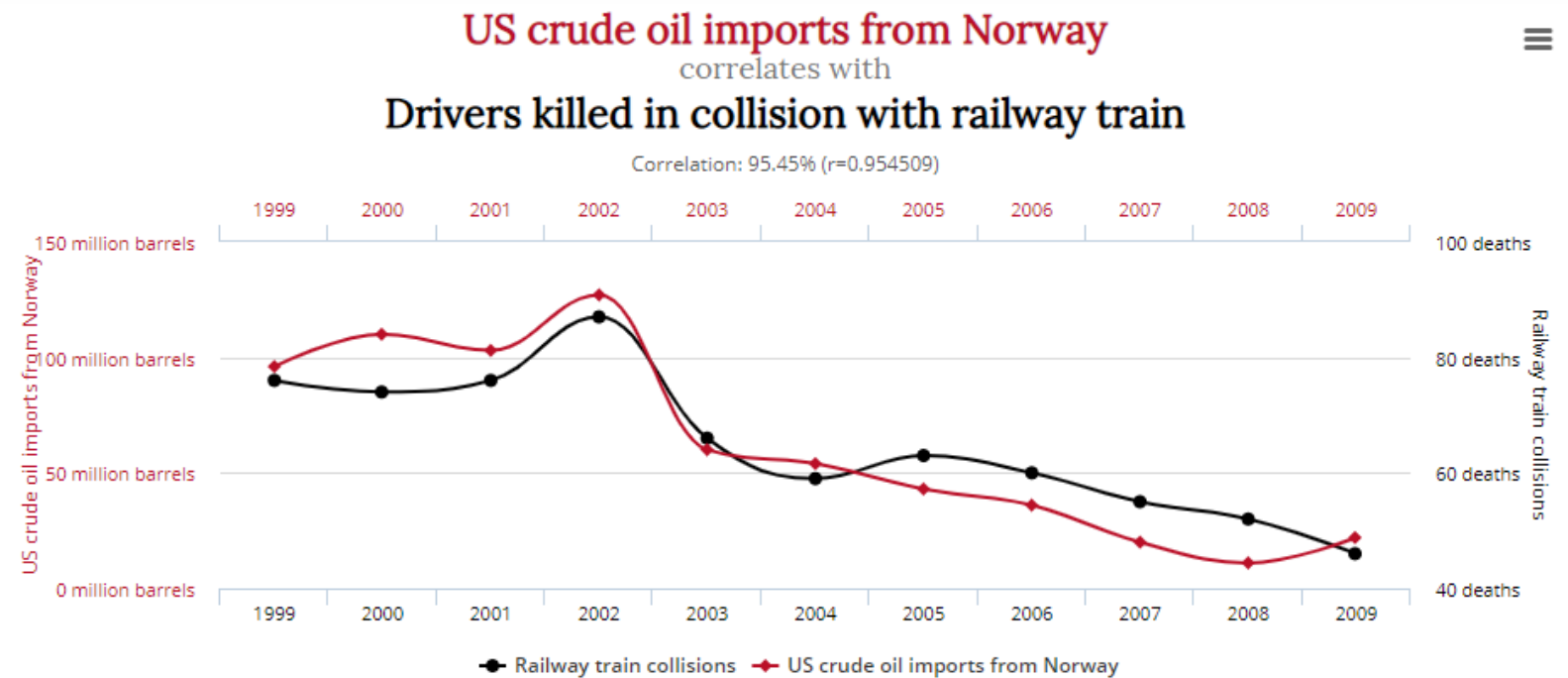- While -1 or 1 means that there is a perfect negative or positive correlation

Example
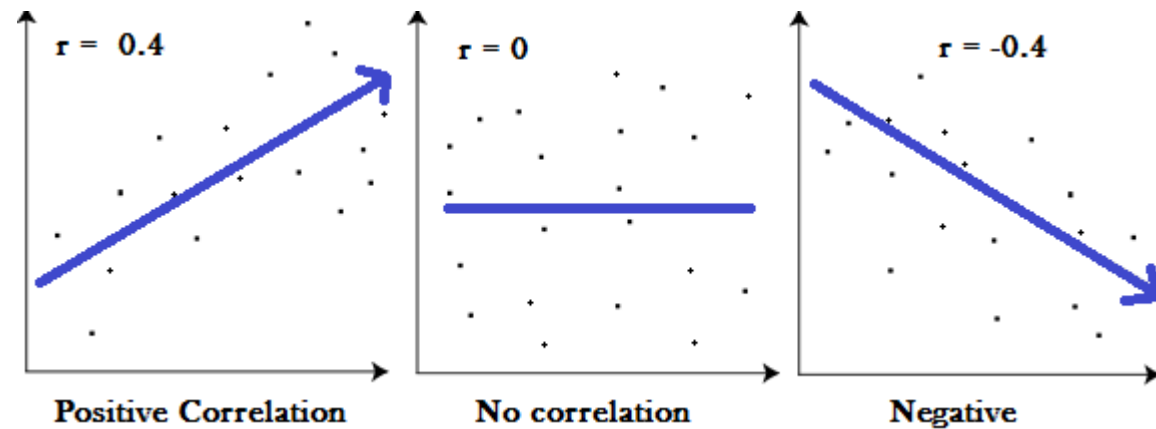
Correlation vs Causation

Causation : The relation between something that happens and the thing that causes it . The first thing that happens is the cause and the second thing is the effect .

Correlation vs Causation

US crude oil imports from Norway correlates with Drivers killed in collision with railway train
Correlation: 95.45% (r=0.954509)

The Correlation Matrix

Correlation Matrix

Definition

- A correlation matrix is a table showing correlation coefficients between variables.
- Each cell in the table shows the correlation between two variables.
- A correlation matrix is used as a way to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses.

Example

	Always to vote in elections	Never to try to evade taxes	Always to obey laws	Keep watch on action of govt
Always to vote in elections	1.00	.94	.94	.94
Never to try to evade taxes	.94	1.00	.97	.95
Always to obey laws	.94	.97	1.00	.96
Keep watch on action of govt	.94	.95	.96	1.00

Application

- To summarize a large amount of data where the goal is to see patterns.
- To Identify collinearity in the data

Multicollinearity

Multicollinearity

Definition

- Multicollinearity exists whenever two or more of the predictors in a regression model are moderately or highly correlated.

Effects

- Multicollinearity results in a change in the signs as well as in the magnitudes of the partial regression coefficients from one sample to another sample.
- Multicollinearity makes it tedious to assess the relative importance of the independent variables in explaining the variation caused by the dependent variable.

Solution

- Remove highly correlated independent variables by looking at the correlation matrix and VIF

Start-Tech
ACADEMY