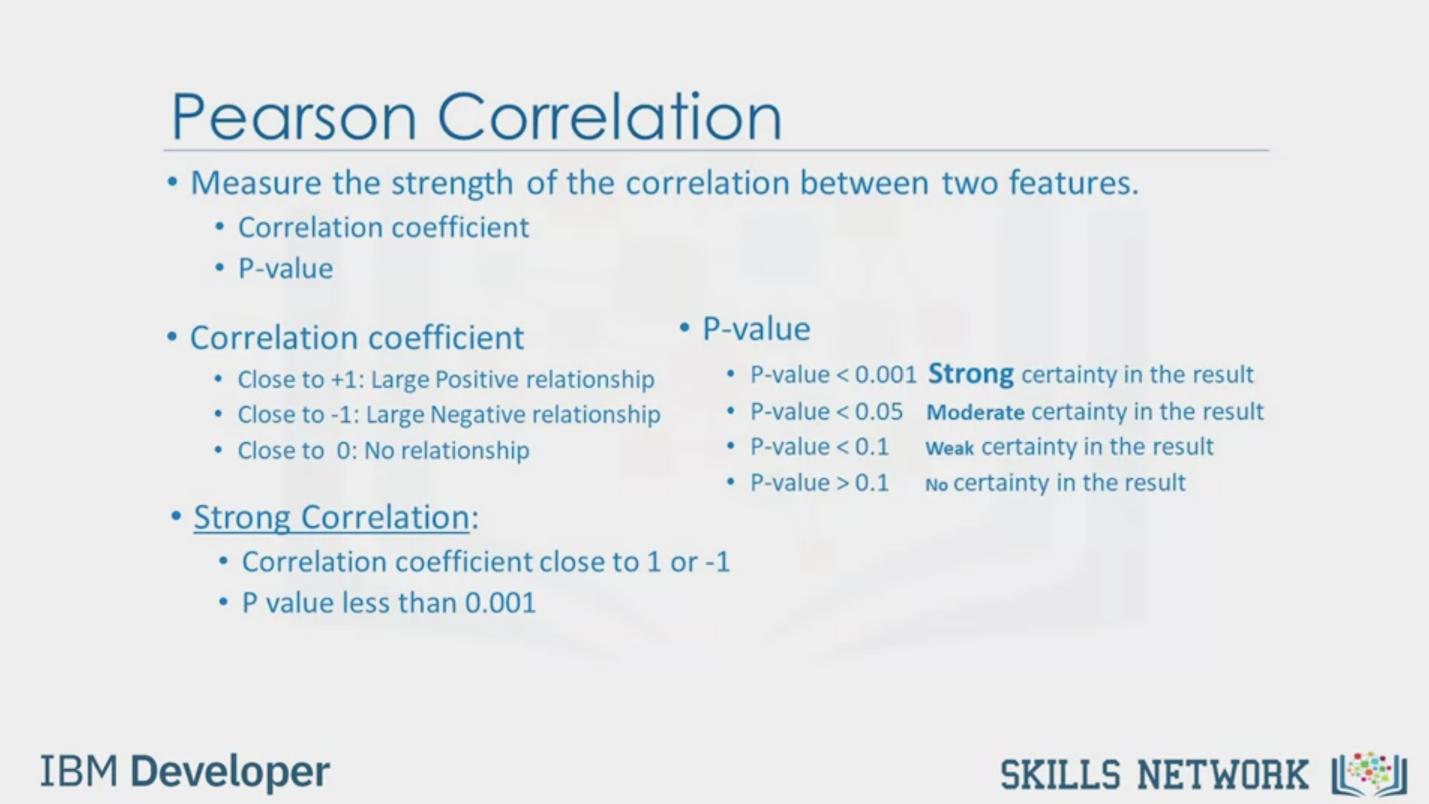


Correlation - Statistics

IBM Developer

SKILLS NETWORK



Pearson Correlation

• Measure the strength of the correlation between two features.

• Correlation coefficient

• p-value

• Correlation coefficient

• Close to +1: Large Positive relationship

• Close to -1: Large Negative relationship

• Close to 0: No relationship

• p-value

• -value <0.001 Strong certainty in the result

• -value < 0.05 Moderate certainty in the result

• -value <0.1 Weak certainty in the result

• -value > 0.1 No certainty in the result

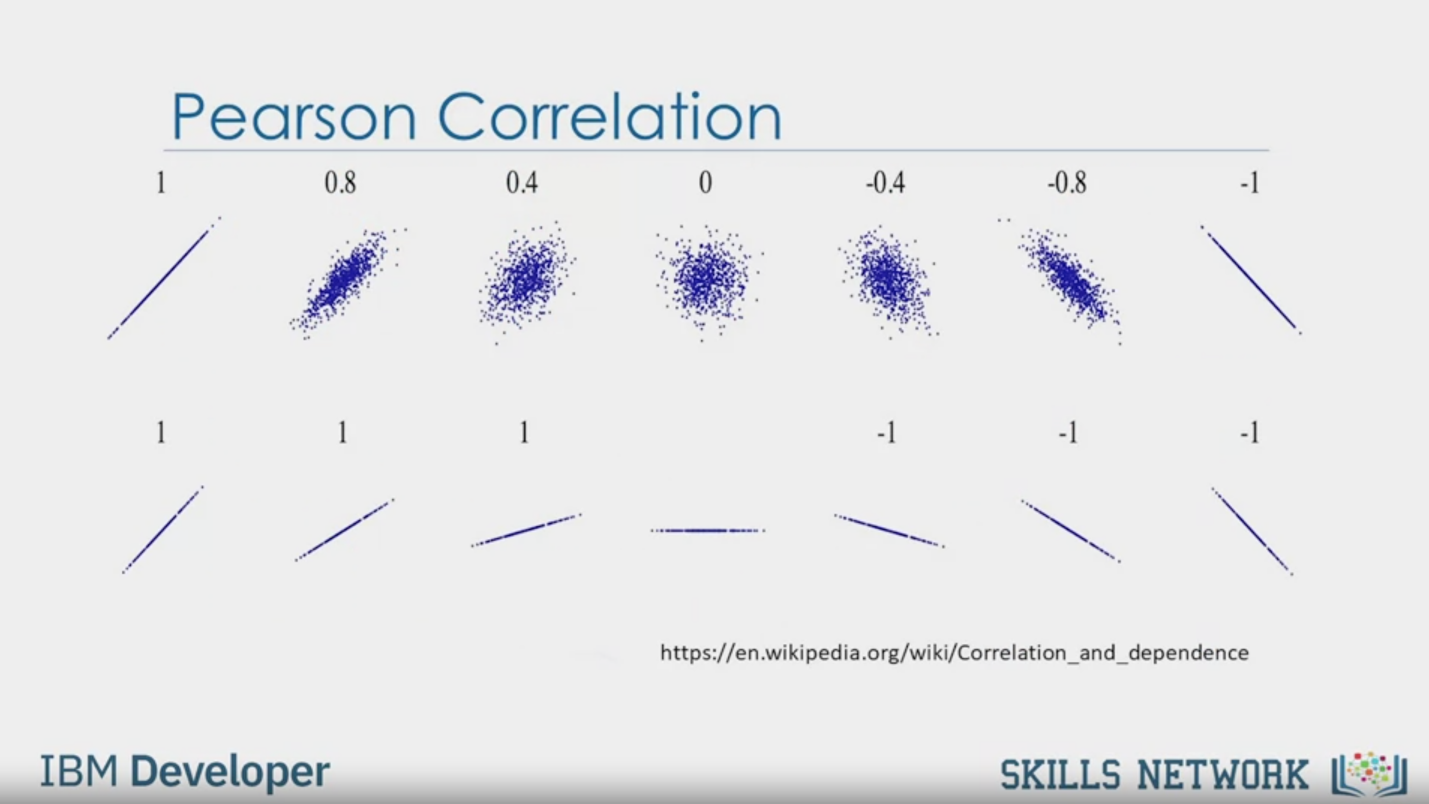
• Strong Correlation:

• Correlation coefficient close to 1 or -1

• P value less than 0.001

IBM Developer

SKILLS NETWORK

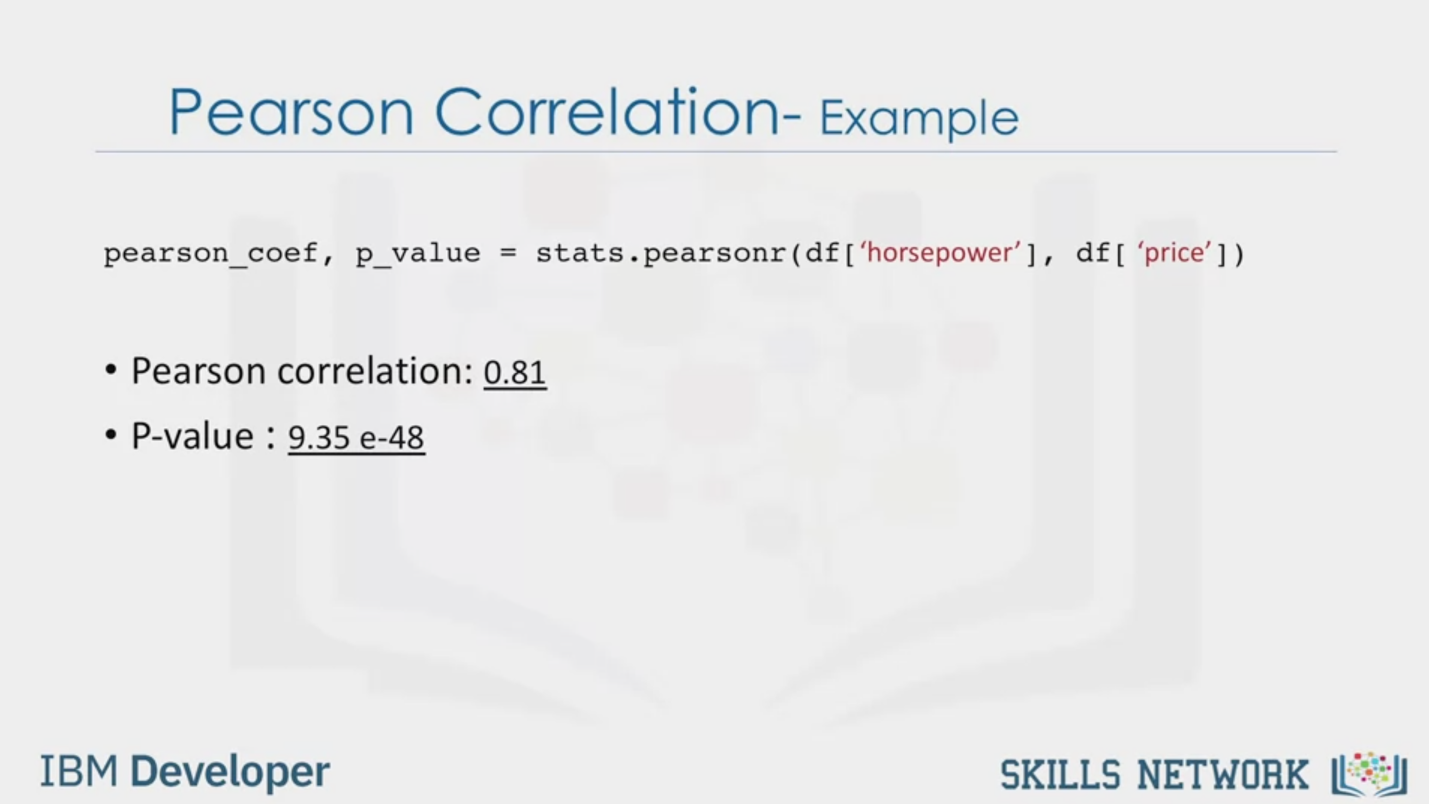


Pearson Correlation

https://en.wikipedia.org/wiki/Correlation\_and\_dependence

IBM Developer

SKILLS NETWORK



Pearson Correlation- Example

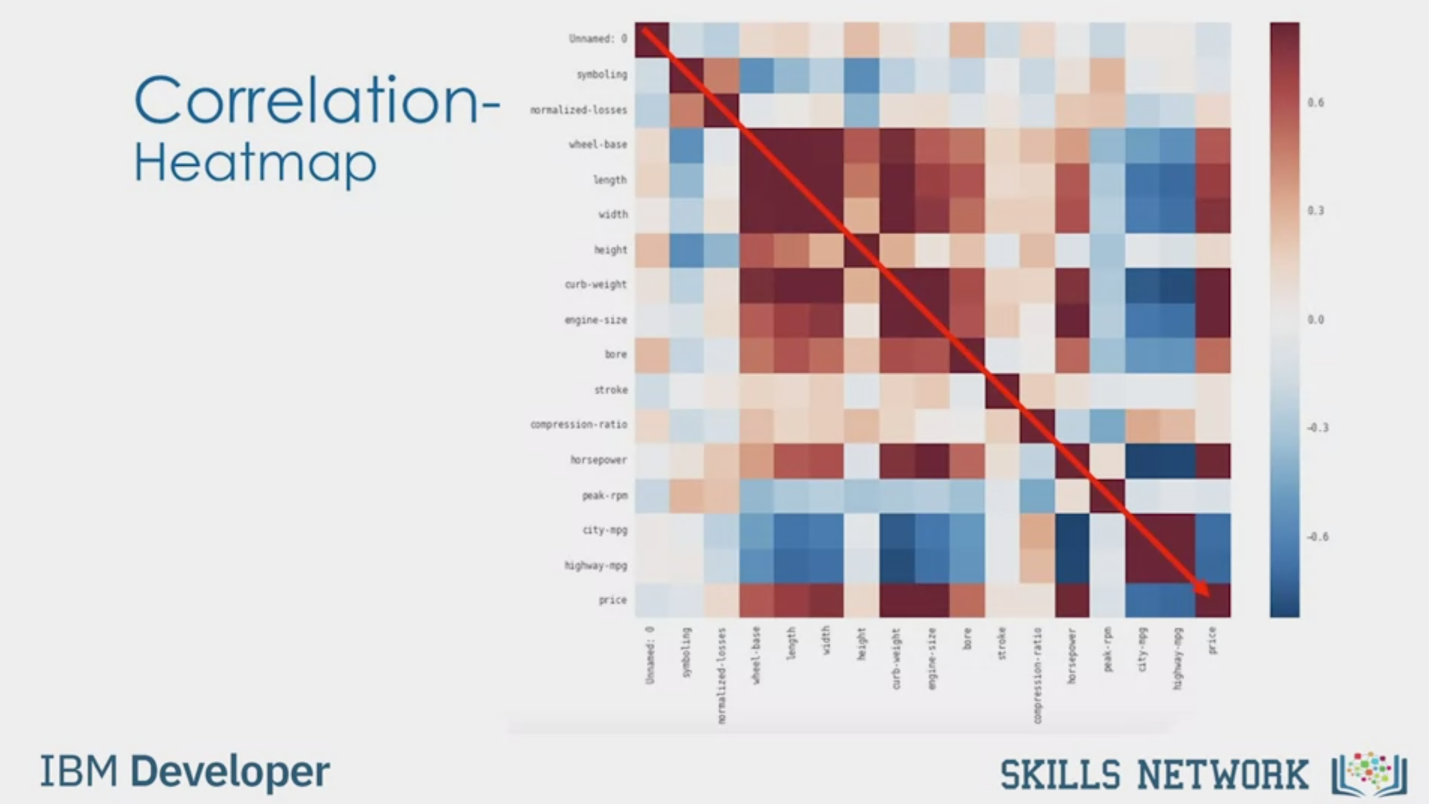
Pearson\_coef, p\_value = stats.pearsonr (df [ 'horsepower'], df ['price'])

• Pearson correlation: 0.81

• -value : 9.35 e-48

IBM Developer

SKILLS NETWORK



Correlation-

Heatmap

IBM Developer

SKILLS NETWORK



SKILLS NETWORK

IBM Developer

In this video, we'll introduce you to various correlations statistical methods.

One way to measure the strength of the correlation between

continuous numerical variable is by using a method called Pearson correlation.

Pearson correlation method will give you

two values: the correlation coefficient and the P-value.

So how do we interpret these values?

For the correlation coefficient,

a value close to 1 implies a large positive correlation,

while a value close to negative 1 implies a large negative correlation,

and a value close to zero implies no correlation between the variables.

Next, the P-value will tell us how

certain we are about the correlation that we calculated.

For the P-value, a value less than.001

gives us a strong certainty about the correlation coefficient that we calculated.

A value between.001 and.05 gives us moderate certainty.

A value between.05 and.1 will give us a weak certainty.

And a P-value larger than.1 will give us no certainty of correlation at all.

We can say that there is a strong correlation when

the correlation coefficient is close to 1 or negative 1,

and the P-value is less than.001.

The following plot shows data with different correlation values.

In this example, we want to look at the correlation

between the variable's horsepower and car price.

See how easy you can calculate the Pearson correlation using the SI/PI stats package?

We can see that the correlation coefficient is approximately.8, and this is close to 1.

So there is a strong positive correlation.

We can also see that the P-value is very small, much smaller than.001.

And so we can conclude that we are certain about the strong positive correlation.

Taking all variables into account,

we can now create a heatmap that indicates

the correlation between each of the variables with one another.

The color scheme indicates the Pearson correlation coefficient,

indicating the strength of the correlation between two variables.

We can see a diagonal line with a dark red color,

indicating that all the values on this diagonal are highly correlated.

This makes sense because when you look closer,

the values on the diagonal are the correlation of all variables with themselves,

which will be always 1.

This correlation heatmap gives us a good overview

of how the different variables are related to one another and,

most importantly, how these variables are related to price.