# Theoretical principle

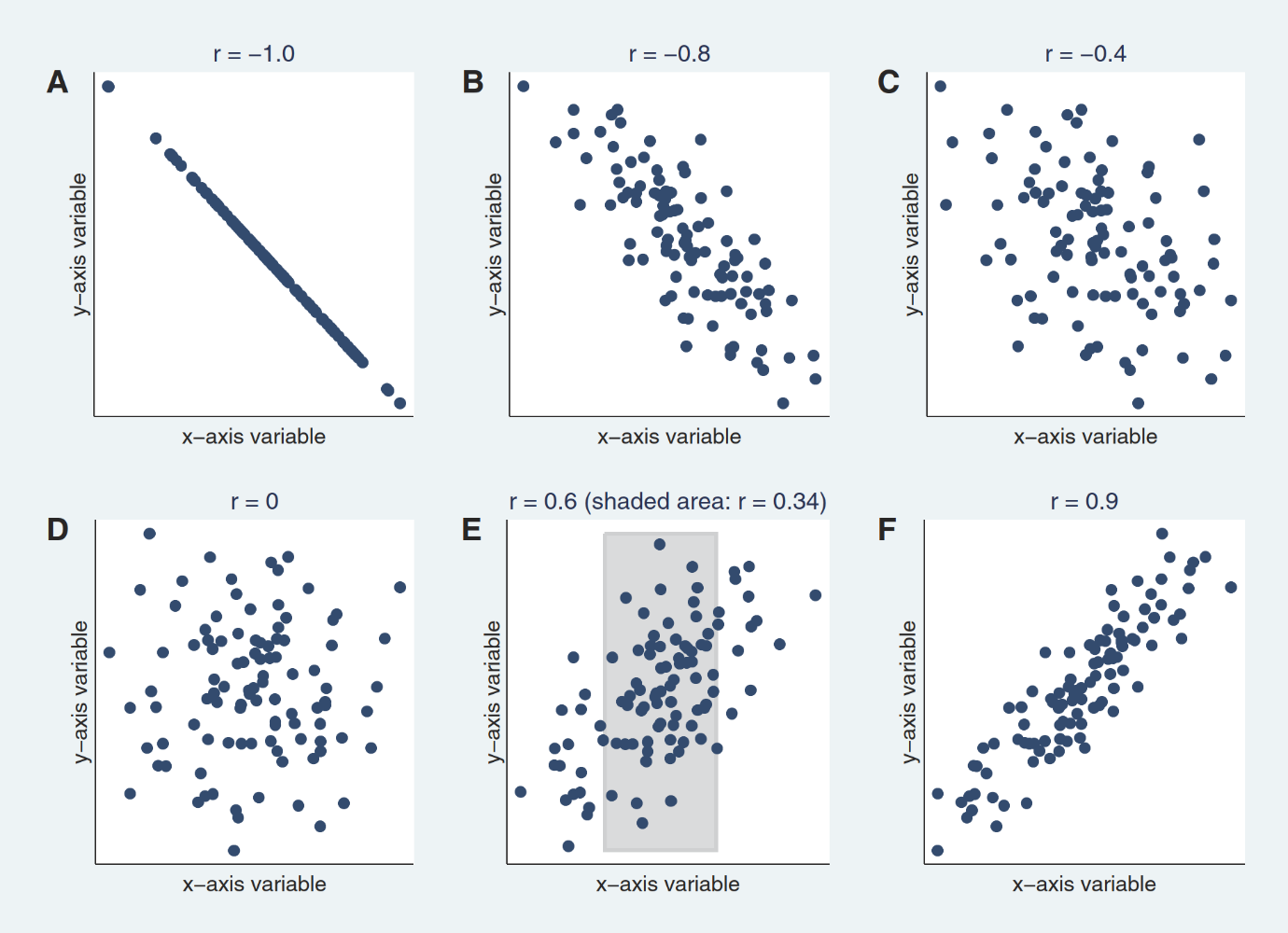
This chapter introduces the basic knowledge of the two most commonly used correlation coefficients – the Pearson coefficient and the Spearman coefficient – required to comprehend the correlation analysis performed in **chapter 4**. We focus on how they should and should not be used and interpreted. **Section XY** covers the fundamentals of seasonal decomposition, a concept we used to decompose the restaurant revenue time series.

## Correlation coefficients

Correlation is a measure of a monotonic relationship between two variables in a correlated data, where the increase of the value of one variable tend to result in either an increase (positive correlation) or a decrease (negative correlation) of the value of the other one, and vice versa. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

### Peason product-moment correlation

A special case of a monotonic association is a linear relationship between two variables. Most often, the term correlation is used in conjunction with such a linear relationship, known as Pearson product-moment correlation, commonly abbreviated as *r*. This coefficient is a dimensionless measure and ranges from -1 to 1. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

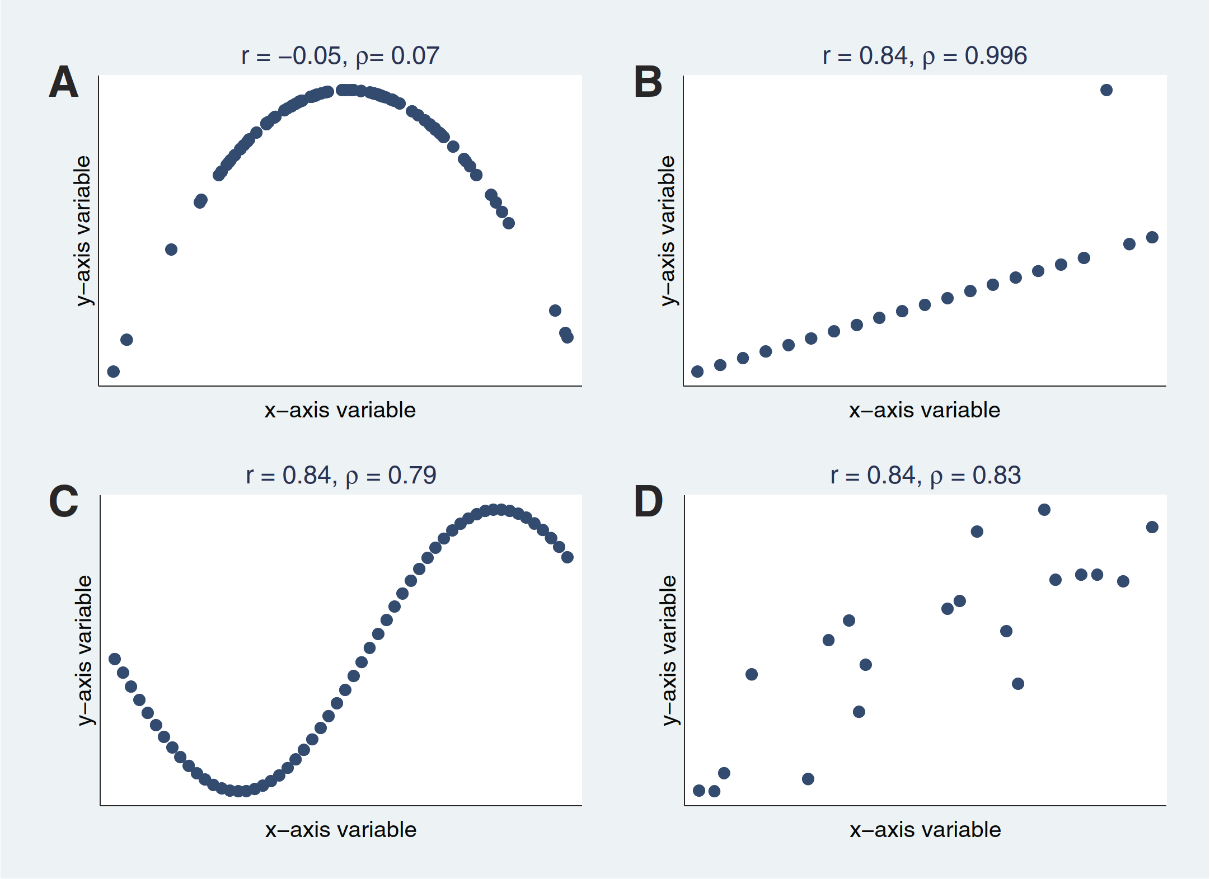
The **figure** below depicts scatterplots of sample data with different Pearson correlation coefficients.

**Figure XY** A illustrates a perfect correlation of -1. A perfect correlation of -1 or 1 implies that all the data points lie exactly on a straight line. In **Figure XY** B and F, the scatterplot approaches a straight line as the coefficient tends towards -1 or 1, whereas in **Figure XY** D there is no linear relationship, as the coefficient is 0. **Figure XY** E displays that the correlation depends on the range of the assessed value, a wider range leans towards higher correlation than the smaller range in the shaded area. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

### Spearman rank correlation

In contrast to a Pearson correlation, a Spearman correlation – generally abbreviated as *ρ* (rho) or *rs* – can be used to analyse nonlinear monotonic relationships. Furthermore, it is relatively robust against outliers. The Spearman correlation also ranges from -1 to 1, whereas *ρ* = 0 implies that there is no association, while *ρ* = -1 or 1 implicate a perfect correlation. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

### Interpretation of the correlation coefficients

The scatterplots in the following **figure XY** illustrate the two correlation methods – Pearson and Spearman – on a sample dataset. Note, that the correlation coefficient should always be assessed by a visual representation of the data. For example, in **figure XY** A, both coefficients are close to 0, which connotes that there is no association between the x-axis and y-axis variables, when in fact, the plot suggests a strong quadratic relationship. Another interesting observation is, that despite the same Pearson correlation coefficient values *r* in **figures XY B through D**, the data is quite different in each of the panels. **Figure XY B** reveals, on the one hand, the robustness of the Spearman coefficient against outliers and on the other hand, its notable influence on the Pearson coefficient. In **figure XY** C, a sinusoid relationship – neither linear nor monotonic – is depicted, both correlation methods are unable to capture it. This can be further observed in **figure XY D**.

Over the course of years, several threshold values to translate a correlation coefficient into descriptors such as “weak”, “moderate” or “strong” relationship – which are arbitrary and inconsistent – have been proposed. While most researchers would agree that a correlation less than 0.1 indicates a negligible and one greater than 0.9 a strong relationship, values in between are disputable and therefore should be interpreted within the context of the posed research question. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

## Time series decomposition

Time series decomposition is a method that splits a time series, which may exhibit a variety of patterns into several components, each representing an underlying pattern category, “trend”, “seasonality”, and “residual”. **(Source:** [**https://otexts.com/fpp2/decomposition.html**](https://otexts.com/fpp2/decomposition.html)**,** [**https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2**](https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2)**)** This is often employed to help improve understanding of the time series or to improve forecast accuracy. **(**[**https://otexts.com/fpp2/decomposition.html**](https://otexts.com/fpp2/decomposition.html)**)**

The decomposed pattern components are defined as follows:

* **Trend** describes whether the time series is decreasing, increasing or constant over time. (**Source:** [**https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2**](https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2)) It does not have to be linear. Sometimes we refer to a trend as “changing direction” when it might go from an increasing to a decreasing trend. **(**[**https://otexts.com/fpp2/tspatterns.html**](https://otexts.com/fpp2/tspatterns.html)**, paraphrase this)**
* **Seasonality** describes the periodic signal in the time series. (**Source:** [**https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2**](https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2)) This pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality has always a fixed and known frequency. **(**[**https://otexts.com/fpp2/tspatterns.html**](https://otexts.com/fpp2/tspatterns.html)**, paraphrase this)**
* **Residual** is what remains behind the separation of seasonality and trend from the time series. It is the variability in the data that cannot be explained. (**Source:** [**https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2**](https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2))

A time series can be considered as a combination of these components, either additively or multiplicatively. **(Source:** [**https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/**](https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/)**, paraphrase this)**

An additive decomposition model is defined as

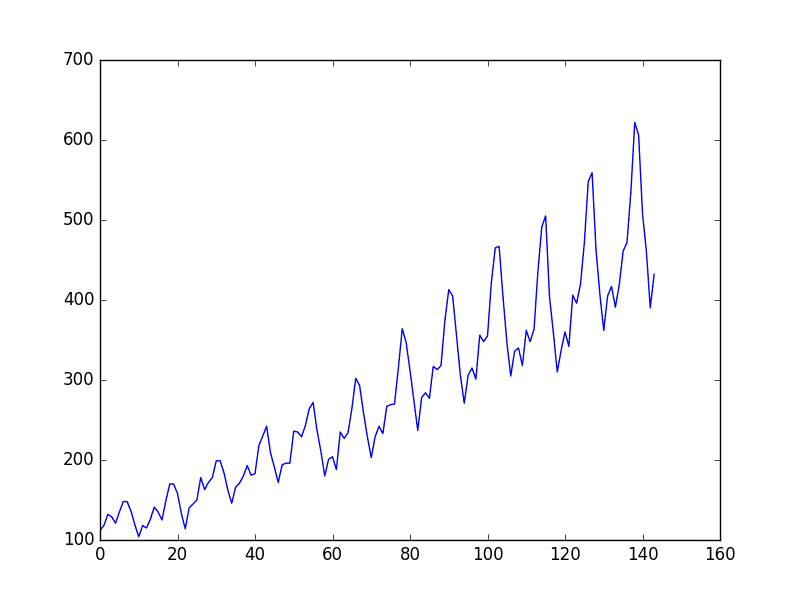
,

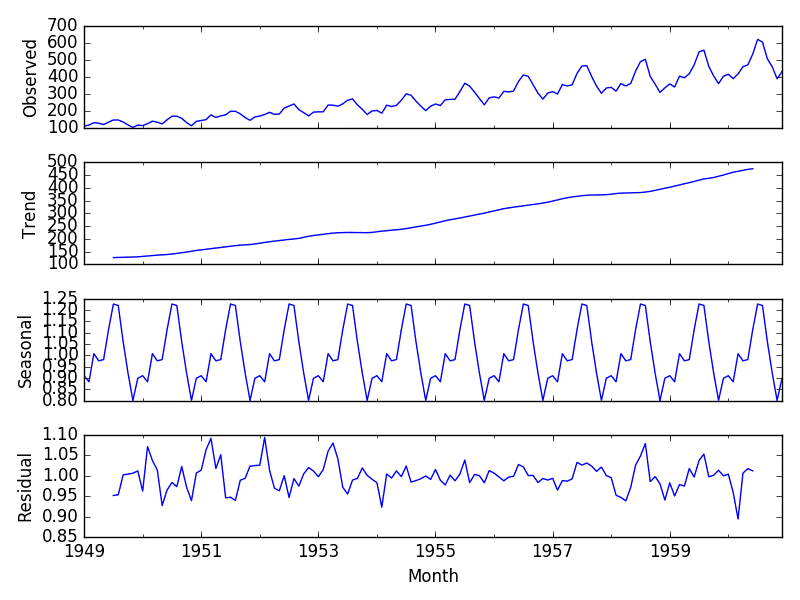
where is the data; , the seasonal component; , the trend component; and , the residual component at period . Alternatively, a multiplicative decomposition is formulated as (**Source:** [**https://otexts.com/fpp2/components.html**](https://otexts.com/fpp2/components.html))

.

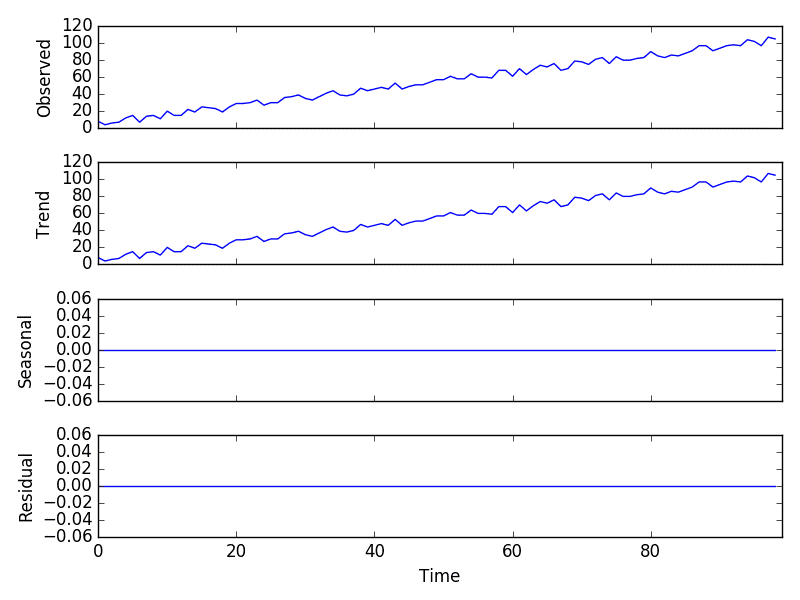
To identify whether the problem is additive or multiplicative, a review of a plot of the time series can be regarded as a good starting point. **(Source:** [**https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/**](https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/)**)** A rule of thumb for selecting the right model is to see if the trend and seasonal variation are relatively constant over time, i.e., linear. When this is the case, an additive model can be chosen. Otherwise, if the trend and seasonal variation increase or decrease over time, a multiplicative decomposition shall be used. (**Source:** [**https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2**](https://towardsdatascience.com/time-series-decomposition-in-python-8acac385a5b2))

### Example on a real-world dataset

For better understanding, let us look at a real-world dataset **(see figure XY)**, which describes the total number of airline passengers from 1949 to 1960. The horizontal axis represents the number of monthly observations during that period, the vertical axis, the number of airline passengers in thousands. **(Source:** [**https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/**](https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/)**)**

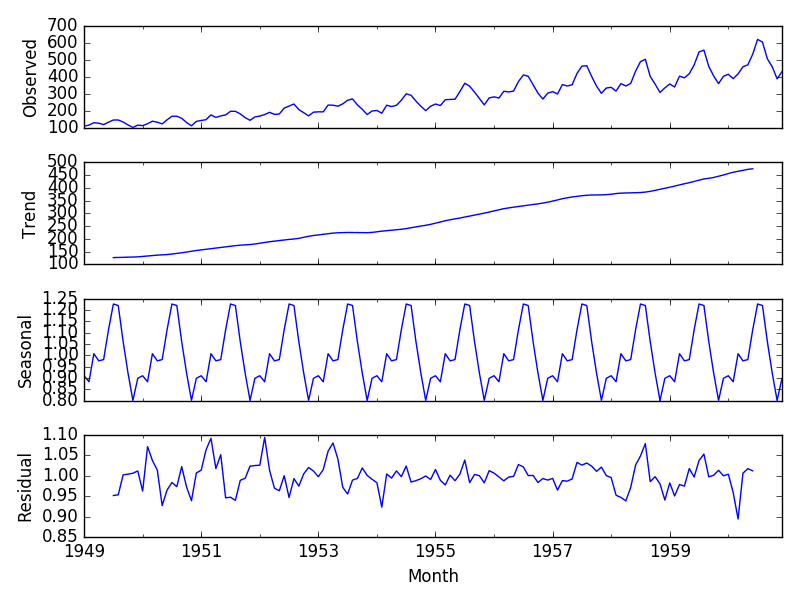
The line plot may suggest a linear trend. A seasonality can also be observed; however, the amplitude appears to be increasing, indicating a multiplicative problem. Hence, a multiplicative decomposition is applied as shown in **figure XY**. **(Source:** [**https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/**](https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/)**)**

There are also datasets for which a naïve or classical decomposition fails, as illustrated in **figure XY**, where it was not able to separate the noise from the linear trend. For these scenarios, caution and scepticism is needed. **(Source:** [**https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/**](https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/)**)**



Time series data can exhibit a variety of patterns, and it is often helpful to split a time series into several components, each representing an underlying pattern category.(<https://otexts.com/fpp2/decomposition.html>)

In describing these time series, we have used words such as “trend” and “seasonal” which need to be defined more carefully.(https://otexts.com/fpp2/tspatterns.html)

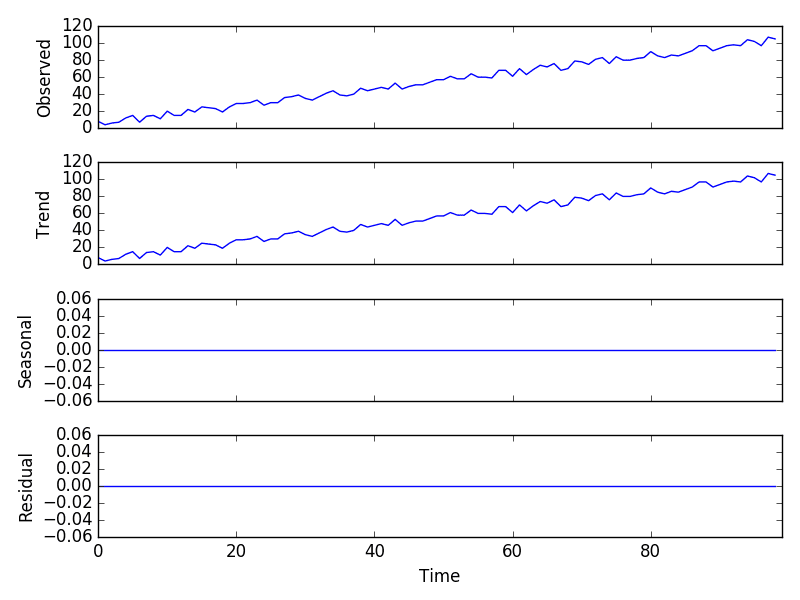


(https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality)

A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we will refer to a trend as “changing direction”, when it might go from an increasing trend to a decreasing trend. (https://otexts.com/fpp2/tspatterns.html) There is a visible rising trend in the example data shown in Figure [number]

A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known frequency. (https://otexts.com/fpp2/tspatterns.html) There is an observable seasonality in the example data shown in Figure [number]

Residual is the resulting data after extracting the seasonality and trend from the observed data.



(https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality)

There are also data sets in which there is only a visible trend but no seasonality as seen in Figure [number]. Or it could be the other way around. For this reason, we have to look at our data and analyze them the same way to conclude which combination of these components to use for our evaluation.