**TIME SERIES DATA:**

A time series is a set of observations on the values that a variable takes at different times.

Such data may be collected at regular time intervals, such as, monthly (eg. CPI),weekly (eg. Money supply), quarterly (eg. GDP) or annually (eg. Government Budget).

Time series are used in statistics, econometrics, mathematical finance, weather forecasting, earthquake prediction and many other applications.

**UNIVARIATE TIME SERIES:**

A “Univariate time series” refers to a time series that consists of single observations recorded over regular time intervals.

Examples: Monthly returns data of a stock.

**CROSS- SECTIONAL DATA:**

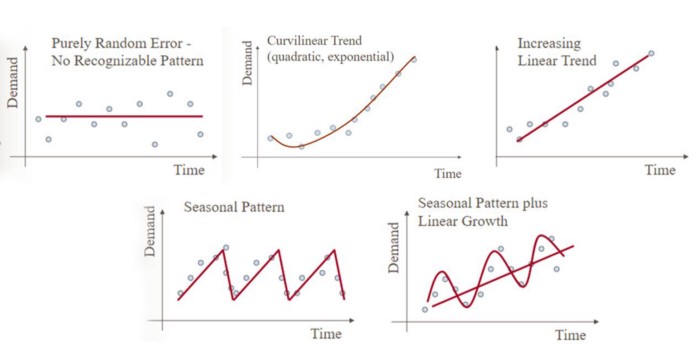
Such a type of data is collected by observing many subjects (such as individuals, firms, countries,or regions) at the same point of time or during the same time period.

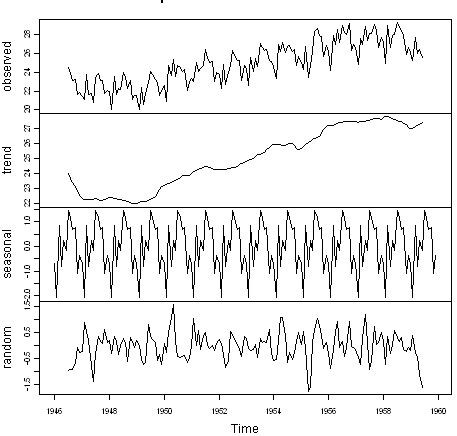
Example: Suppose an analyst wants to know the number of cars a household has bought in the past year. To do so, he collects data on a sample of, say, 500 families from the population and notes the data on how many cars they have bought in the past year.

This “cross-sectional” sample provides a glimpse of the population for that duration.

**PATTERNS EMERGING IN THE TIME SERIES DATA:**

Depending on the frequency of the data (hourly, daily, weekly, monthly, quarterly, annually, etc) different patterns emerge in the data set which forms the component to be modeled.

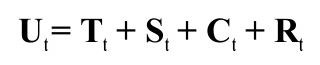
Sometimes the time series may just be increasing or decreasing over time with a constant slope or there may be patterns around the increasing slope. 



A Decomposed time series

**Components of a time series**

A time series can be analyze in detail by breaking down it into its primary components . We call this process as time series decomposition . If you closely look at a time series it is composed of Trend , Seasonality, Cyclic and Residualcomponents.



1. **Trend Component** : The long-term tendency of a series to increase or fall (upward trend or downward trend).

**2. Seasonality Component**: The periodic fluctuation in the time series within a certain period. These fluctuations form a pattern that tends to repeat from one seasonal period to the next one.

**3. Cycles Component** : Long departures from the trend due to factors others than seasonality. Cycles usually occur along a large time interval, and the lengths of time between successive peaks or troughs of a cycle are not necessarily the same.

**4. Irregular movement Component** :The movement left after explaining the trend, seasonal and cyclical movements; random noise or error in a time series.

Using linear regression in the case of continuous output and constant slope:

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It’s used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog). There are two main types:

**Simple regression**

Simple linear regression uses traditional slope-intercept form, where mm and bb are the variables our algorithm will try to “learn” to produce the most accurate predictions. x represents our input data and y represents our prediction.

y=mx+b

**Multivariable regression**

A more complex, multi-variable linear equation might look like this, where ww represents the coefficients, or weights, our model will try to learn.

f(x,y,z)=w1x+w2y+w3z

The variables x,y,z represent the attributes, or distinct pieces of information, we have about each observation. For sales predictions, these attributes might include a company’s advertising spend on radio, TV, and newspapers.

Sales=w1Radio+w2TV+w3News

## Example :[Simple regression](https://ml-cheatsheet.readthedocs.io/en/latest/linear_regression.html" \l "id13)

Let’s say we are given a [dataset](http://www-bcf.usc.edu/~gareth/ISL/Advertising.csv) with the following columns (features): how much a company spends on Radio advertising each year and its annual Sales in terms of units sold. We are trying to develop an equation that will let us to predict units sold based on how much a company spends on radio advertising. The rows (observations) represent companies.

|  |  |  |
| --- | --- | --- |
| **Company** | **Radio ($)** | **Sales** |
| Amazon | 37.8 | 22.1 |
| Google | 39.3 | 10.4 |
| Facebook | 45.9 | 18.3 |
| Apple | 41.3 | 18.5 |

### [Making predictions](https://ml-cheatsheet.readthedocs.io/en/latest/linear_regression.html#id14) :

Our prediction function outputs an estimate of sales given a company’s radio advertising spend and our current values for Weight*and*Bias*.*

Sales=Weight⋅Radio+BiasSales=Weight⋅Radio+Bias

Weight

the coefficient for the Radio independent variable. In machine learning we call coefficients weights*.*

Radio

the independent variable. In machine learning we call these variables features.

Bias

the intercept where our line intercepts the y-axis. In machine learning we can call intercepts bias. Bias offsets all predictions that we make.

Our algorithm will try to learn the correct values for Weight and Bias. By the end of our training, our equation will approximate the line of best fit*.*

