TIME SERIES PROJECT

Submitted by:

Surabhi Grover-1916039

Ankita Yadav-1916064

WHAT IS A TIME SERIES?

A **time series** is a series of [data points](https://en.wikipedia.org/wiki/Data_point) indexed (or listed or graphed) in time order. Most commonly, a time series is a [sequence](https://en.wikipedia.org/wiki/Sequence) taken at successive equally spaced points in time. Thus it is a sequence of [discrete-time](https://en.wikipedia.org/wiki/Discrete-time) data.

Examples of time series are heights of ocean [tides](https://en.wikipedia.org/wiki/Tides), counts of [sunspots](https://en.wikipedia.org/wiki/Sunspots), and the daily closing value of the Dow Jones Industrial Average.

WHERE DO WE USE TIME SERIES?

Time series are very frequently plotted via [line charts](https://en.wikipedia.org/wiki/Line_chart). Time series are used in [statistics](https://en.wikipedia.org/wiki/Statistics), [signal processing](https://en.wikipedia.org/wiki/Signal_processing), [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), [econometrics](https://en.wikipedia.org/wiki/Econometrics), [mathematical finance](https://en.wikipedia.org/wiki/Mathematical_finance), [weather forecasting](https://en.wikipedia.org/wiki/Weather_forecasting), [earthquake prediction](https://en.wikipedia.org/wiki/Earthquake_prediction), [electroencephalography](https://en.wikipedia.org/wiki/Electroencephalography), [control engineering](https://en.wikipedia.org/wiki/Control_engineering), [astronomy](https://en.wikipedia.org/wiki/Astronomy), [communications engineering](https://en.wikipedia.org/wiki/Communications_engineering), and largely in any domain of applied [science](https://en.wikipedia.org/wiki/Applied_science) and [engineering](https://en.wikipedia.org/wiki/Engineering) which involves [temporal](https://en.wikipedia.org/wiki/Time) measurements.

TIME SERIES ANALYSIS:

**Time series *analysis*** comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

**Time series *forecasting*** is the use of a [model](https://en.wikipedia.org/wiki/Model_(abstract)) to predict future values based on previously observed values. While [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time. [Interrupted time series](https://en.wikipedia.org/wiki/Interrupted_time_series) analysis is the analysis of interventions on a single time series.

**COMPONENTS OF A TIME SERIES:**

* Secular Trends (or simply trends)

The secular trend is the main component of a time series which results from long term effects of socio-economic and political factors. This trend may show the growth or decline in a time series over a long period. This is the type of tendency which continues to persist for a very long period. Prices and export and import data, for example, reflect obviously increasing tendencies over time.

* Seasonal Movements.

These are short term movements occurring in data due to seasonal factors. The short term is generally considered as a period in which changes occur in a time series with variations in weather or festivities. For example,  it is commonly observed that the consumption of ice-cream during summer is generally high and hence an ice-cream dealer's sales would be higher in some months of the year while relatively lower during winter months. Employment, output, exports, etc., are subject to change due to variations in weather. Similarly, the sale of garments, umbrellas, greeting cards and fire-works are subject to large variations during festivals like Valentine’s Day, Eid, Christmas, New Year's, etc. These types of variations in a time series are isolated only when the series is provided biannually, quarterly or monthly.

* Cyclical Movements.

These are long term oscillations occurring in a time series. These oscillations are mostly observed in economics data and the periods of such oscillations are generally extended from five to twelve years or more. These oscillations are associated with the well known business cycles. These cyclic movements can be studied provided a long series of measurements, free from irregular fluctuations, is available.

* Irregular Fluctuations.

These are sudden changes occurring in a time series which are unlikely to be repeated. They are components of a time series which cannot be explained by trends, seasonal or cyclic movements. These variations are sometimes called residual or random components. These variations, though accidental in nature, can cause a continual change in the trends, seasonal and cyclical oscillations during the forthcoming period. Floods, fires, earthquakes, revolutions, epidemics, strikes etc., are the root causes of such irregularities.

**VARIOUS MODELS OF THE TIME SERIES:**

**Additive models**

The additive model is based on the assumption that the different components affected the time series additively.

http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/Hseasonal/images/dataComponents.gif

For monthly data, an additive model assumes that the difference between the January and July values is approximately the same each year. In other words, the **amplitude** of the seasonal effect is the same each year.

The model similarly assumes that the residuals are roughly the same size throughout the series -- they are a random component that adds on to the other components in the same way at all parts of the series.

**Multiplicative models**

In many time series involving **quantities** (e.g. money, wheat production, ...), the absolute differences in the values are of less interest and importance than the percentage changes.

For example, in seasonal data, it might be more useful to model that the July value is the same **proportion** higher than the January value in each year, rather than assuming that their difference is constant. Assuming that the seasonal and other effects act proportionally on the series is equivalent to a **multiplicative model**,

http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/Hmultiplicative/images/multModel.gif

Fortunately, multiplicative models are equally easy to fit to data as additive models! The trick to fitting a multiplicative model is to take logarithms of both sides of the model,

http://www-ist.massey.ac.nz/dstirlin/CAST/CAST/Hmultiplicative/images/multLogModel.gif

After taking logarithms (either natural logarithms or to base 10), the four components of the time series again act additively.

|  |
| --- |
| **To fit a multiplicative model, take logarithms of the data, then analyse the log data as before.** |

**TREND COMPONENT OF A TIME SERIES**

The trend is the long term pattern of a time series. A trend can be positive or negative depending on whether the time series exhibit an increasing long term pattern or a decreasing long term pattern. If a time series does not show the increasing or decreasing pattern then the time series is stationary in the mean.

For example, an upward tendency would be seen in data pertaining to population, agricultural production, currency in circulation etc., while a downward tendency will be noticed in data of births and deaths, epidemics etc., as a result of advancement in medical sciences, better medical facilities, literacy and higher standard of living.

Trend is the general, smooth, long term, average tendency. It is not necessary that the increase or decline should be in the same direction throughout the given period. It may be possible that different tendencies of increase, decrease or stability are observed in different sections of time. Such tendencies are the result of the forces which are, more or less, constant for long time or which change very gradually and continuously over a long period of time such as the change in the population, taste, habits and customs of the people in a society and so on. They operate in an evolutionary manner and do not reflect sudden changes. For example, the effect of the population increase over a long period of time on the expansion of various sectors like agriculture, industry, education, textiles, etc., is a continuous but a gradual process. Similarly, the growth or decline in a no. of economic time series is the interaction of forces like advances in production technology, large scale production, improved marketing management and business organization, the invention and discovery of new natural resources and the exhaustion of the existing resources and so on- all of which are gradual processes.

**MEASUREMENT OF TREND**

Trend can be studied and measured by the following methods:

1. Graphical ( free hand curve fitting) method
2. Method of semi averages
3. Method of curve fitting by principle of least squares
4. Method of moving averages

Method 1)

Graphical method:

A free hand smooth curve obtained on plotting the values yt against ‘t’ enables us to firm an idea about the general ‘trend’ of the series.

This method does not involve any complex mathematical techniques and can be used to describe all types of trend, linear and non-linear. Thus, simplicity and flexibility are strong points of this method.

Consider the following data of number of no. of normal births in each hour in four hospital series, transformed to y=square root of number of births.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr.no.** |  |  | **SQRT(Births)=y** |
| 1 | **AM** | 12 | 74.46 |
| 2 |  | 1 | 74.58 |
| 3 |  | 2 | 76.14 |
| 4 |  | 3 | 78.92 |
| 5 |  | 4 | 78.31 |
| 6 |  | 5 | 79.72 |
| 7 |  | 6 | 77.64 |
| 8 |  | 7 | 75.71 |
| 9 |  | 8 | 78.9 |
| 10 |  | 9 | 78.61 |
| 11 |  | 10 | 76.79 |
| 12 |  | 11 | 74.34 |
| 13 | **PM** | 12 | 72.57 |
| 14 |  | 1 | 71.18 |
| 15 |  | 2 | 70.75 |
| 16 |  | 3 | 66.52 |
| 17 |  | 4 | 70.57 |
| 18 |  | 5 | 68.21 |
| 19 |  | 6 | 68.83 |
| 20 |  | 7 | 65.91 |
| 21 |  | 8 | 70.2 |
| 22 |  | 9 | 70.88 |
| 23 |  | 10 | 70.09 |
| 24 |  | 11 | 72.17 |

**Conclusion based on the above graph:**

The figure shows the number of normal human births in each hour in four hospital series (transformed to y= square root of number of births. We can clearly see that the number of births in the first half of the day is comparatively more than the births in the second half. It clearly shows a declining trend till 20th hour and starts to increase afterwards.

**METHOD 2**

**Method of semi-averages:**

In this method the whole data is divided into two parts with respect to time. In case of odd number of years the two parts are obtained by omitting the value corresponding to the middle year. Next we compute the arithmetic mean for each part and plot these two averages against the mid-values of the respective time periods covered by each part. The line obtained on joining these two points is the required trend line and may be extended both ways to estimate intermediate or future values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr.no.** |  |  | **SQRT(Births)=y** | **Semi average** |
| 1 | **AM** | 12 | 74.46 |  |
| 2 |  | 1 | 74.58 |  |
| 3 |  | 2 | 76.14 |  |
| 4 |  | 3 | 78.92 |  |
| 5 |  | 4 | 78.31 |  |
| 6 |  | 5 | 79.72 | 77.01 |
| 7 |  | 6 | 77.64 |  |
| 8 |  | 7 | 75.71 |  |
| 9 |  | 8 | 78.9 |  |
| 10 |  | 9 | 78.61 |  |
| 11 |  | 10 | 76.79 |  |
| 12 |  | 11 | 74.34 |  |
| 13 | **PM** | 12 | 72.57 |  |
| 14 |  | 1 | 71.18 |  |
| 15 |  | 2 | 70.75 |  |
| 16 |  | 3 | 66.52 |  |
| 17 |  | 4 | 70.57 |  |
| 18 |  | 5 | 68.21 | 69.8233 |
| 19 |  | 6 | 68.83 |  |
| 20 |  | 7 | 65.91 |  |
| 21 |  | 8 | 70.2 |  |
| 22 |  | 9 | 70.88 |  |
| 23 |  | 10 | 70.09 |  |
| 24 |  | 11 | 72.17 |  |

Conclusion based on above graph:

Here series 1 represents the time series values and series 2 represents the semi averages (along with a trend line by joining the two points of the averages the data).

**METHOD 3**

**Method of curve fitting by principle of least squares:**

The method of least squares is the most popular and widely used method of fitting the mathematical functions to a given set of data. The method yields very correct results if sufficiently good appraisal of the form of the function to be fitted is obtained either by a scrutiny of the graphical plot of the values over time or by a theoretical understanding of the mechanism of the variable change. An examination of the plotted data provides an adequate basis for deciding upon the type of trend to use.

The various types of curves that may be used to describe the given data in practice are:

(i) A straight line : Yt = a+bt

(ii) A second degree parabola : Yt = a+bt+ct2

(iii) Kth degree polynomial : Yt = a0+a1t+a2t2+…+aktk

(iv) Exponential curves: Yt = abt

logYt = loga +t logb

1. Second degree curve fitted to logarithms:

Yt  = abtct^2

log Yt = loga+t logb+ t2logc = A+Bt+Ct2

(vi) growth curves:

(a) Yt =a+bct ( modified exponential curve)

(b) Yt = ( gompertz curve)

(c) Yt = (logistic curve)

METHOD OF MOVING AVERAGES :

A moving average is a technique to get an overall idea of the [trends](https://www.statisticshowto.datasciencecentral.com/trend-analysis/)in a data set; it is an [average](https://www.statisticshowto.datasciencecentral.com/average/)of any subset of numbers. The moving average is extremely useful for **forecasting long-term trends**. You can calculate it for any period of time. For example, if you have sales data for a twenty-year period, you can calculate a five-year moving average, a four-year moving average, a three-year moving average and so on. [Stock market](http://www.nasdaq.com/) analysts will often use a 50 or 200 day moving average to help them see trends in the stock market and (hopefully) forecast where the stocks are headed.

An average represents the “middling” value of a set of numbers. The moving average is exactly the same, but **the average is calculated several times for several subsets of data.**

**METHOD 4**

**APPROXIMATION TO MOVING AVERAGES:**

The moving average method is quite tedious to apply for the values of p and m and for long series .

**Method of iterates averages:**

By iterated average we mean the simple arithmetic mean repeated or iterated a number of times. An iterated average is characterised by the extent of the average and the number of iterations.

**SPENCER’S 15 POINT FORMULA:**

y0 =

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **SQRT(Births)=y** | **[4]** | **[4]^2** | **[5]** | **[4]^2 [5] {-3,3,4,3,-3}** | **[4]^2 [5] {-3,3,4,3,-3}/320** |
| **AM** | **12** | 74.46 |  |  |  |  |  |
|  | **1** | 74.58 |  |  |  |  |  |
|  | **2** | 76.14 | 304.1 |  |  |  |  |
|  | **3** | 78.92 | 307.95 |  |  |  |  |
|  | **4** | 78.31 | 313.09 | 1239.73 |  |  |  |
|  | **5** | 79.72 | 314.59 | 1247.01 |  |  |  |
|  | **6** | 77.64 | 311.38 | 1251.03 | 6230.79 |  |  |
|  | **7** | 75.71 | 311.97 | 1248.8 | 6232.54 |  |  |
|  | **8** | 78.9 | 310.86 | 1244.22 | 6217.35 | 25037.04 | 78.24075 |
|  | **9** | 78.61 | 310.01 | 1241.48 | 6182.16 | 24914.58 | 77.85806 |
|  | **10** | 76.79 | 308.64 | 1231.82 | 6128.03 | 24699.71 | 77.18659 |
|  | **11** | 74.34 | 302.31 | 1215.84 | 6050.86 | 24361.72 | 76.13038 |
| **PM** | **12** | 72.57 | 294.88 | 1194.67 | 5953.14 | 23897.91 | 74.68097 |
|  | **1** | 71.18 | 288.84 | 1167.05 | 5846.25 | 23367.69 | 73.02403 |
|  | **2** | 70.75 | 281.02 | 1143.76 | 5740.63 | 22852.45 | 71.41391 |
|  | **3** | 66.52 | 279.02 | 1124.93 | 5648.68 | 22419.28 | 70.06025 |
|  | **4** | 70.57 | 276.05 | 1110.22 | 5578.48 | 22114.15 | 69.10672 |
|  | **5** | 68.21 | 274.13 | 1102.72 | 5531.34 | 21917.25 | 68.49141 |
|  | **6** | 68.83 | 273.52 | 1096.85 | 5505.98 |  |  |
|  | **7** | 65.91 | 273.15 | 1096.62 | 5505.15 |  |  |
|  | **8** | 70.2 | 275.82 | 1099.57 |  |  |  |
|  | **9** | 70.88 | 277.08 | 1109.39 |  |  |  |
|  | **10** | 70.09 | 283.34 |  |  |  |  |
|  | **11** | 72.17 |  |  |  |  |  |

**SPENCER’S 21 POINT FORMULA:**

y0 =

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **SQRT(Births)=y** | **[5]** | **[5]^2** | **[5]^2 [7]** | **[5]^2 [7] {-1,0,1,2,1,0,-1}** | **[5]^2 [7] {-1,0,1,2,1,0,-1}/350** |
| **AM** | **12** | 74.46 |  |  |  |  |  |
|  | **1** | 74.58 |  |  |  |  |  |
|  | **2** | 76.14 | 382.41 |  |  |  |  |
|  | **3** | 78.92 | 387.67 |  |  |  |  |
|  | **4** | 78.31 | 390.73 | 1941.39 |  |  |  |
|  | **5** | 79.72 | 390.3 | 1949.56 |  |  |  |
|  | **6** | 77.64 | 390.28 | 1949.54 |  |  |  |
|  | **7** | 75.71 | 390.58 | 1943.16 | 13527.33 |  |  |
|  | **8** | 78.9 | 387.65 | 1934.07 | 13445.98 |  |  |
|  | **9** | 78.61 | 384.35 | 1917.28 | 13323.7 |  |  |
|  | **10** | 76.79 | 381.21 | 1892.33 | 13167.46 | 26516.08 | 75.7602286 |
|  | **11** | 74.34 | 373.49 | 1860.04 | 12988.99 | 26061.71 | 74.4620286 |
| **PM** | **12** | 72.57 | 365.63 | 1827.28 | 12794.02 | 25567.22 | 73.0492 |
|  | **1** | 71.18 | 355.36 | 1793.3 | 12604.2 | 25080.75 | 71.6592857 |
|  | **2** | 70.75 | 351.59 | 1764.69 | 12431.77 |  |  |
|  | **3** | 66.52 | 347.23 | 1739.1 | 12290.31 |  |  |
|  | **4** | 70.57 | 344.88 | 1727.46 | 12185.98 |  |  |
|  | **5** | 68.21 | 340.04 | 1719.9 |  |  |  |
|  | **6** | 68.83 | 343.72 | 1718.58 |  |  |  |
|  | **7** | 65.91 | 344.03 | 1722.95 |  |  |  |
|  | **8** | 70.2 | 345.91 |  |  |  |  |
|  | **9** | 70.88 | 349.25 |  |  |  |  |
|  | **10** | 70.09 |  |  |  |  |  |
|  | **11** | 72.17 |  |  |  |  |  |

**Method 5- Using growth curves:**

**Modified exponential curve:**

The modified exponential curve is given by:

Yt=a+bc^t,a>0 (1)

Where yt represents the time series value at the time t and a,b,c are constants called its parameters.

Taking first difference of (1) we get

Δy(t)=y(t+h)-y(t)=bc^t(c^h-1)

Where h is the interval of differencing.

Similarly,

Δy(t-h)=y(t)-y(t-h)=bc^(t-h)(c^h-1)

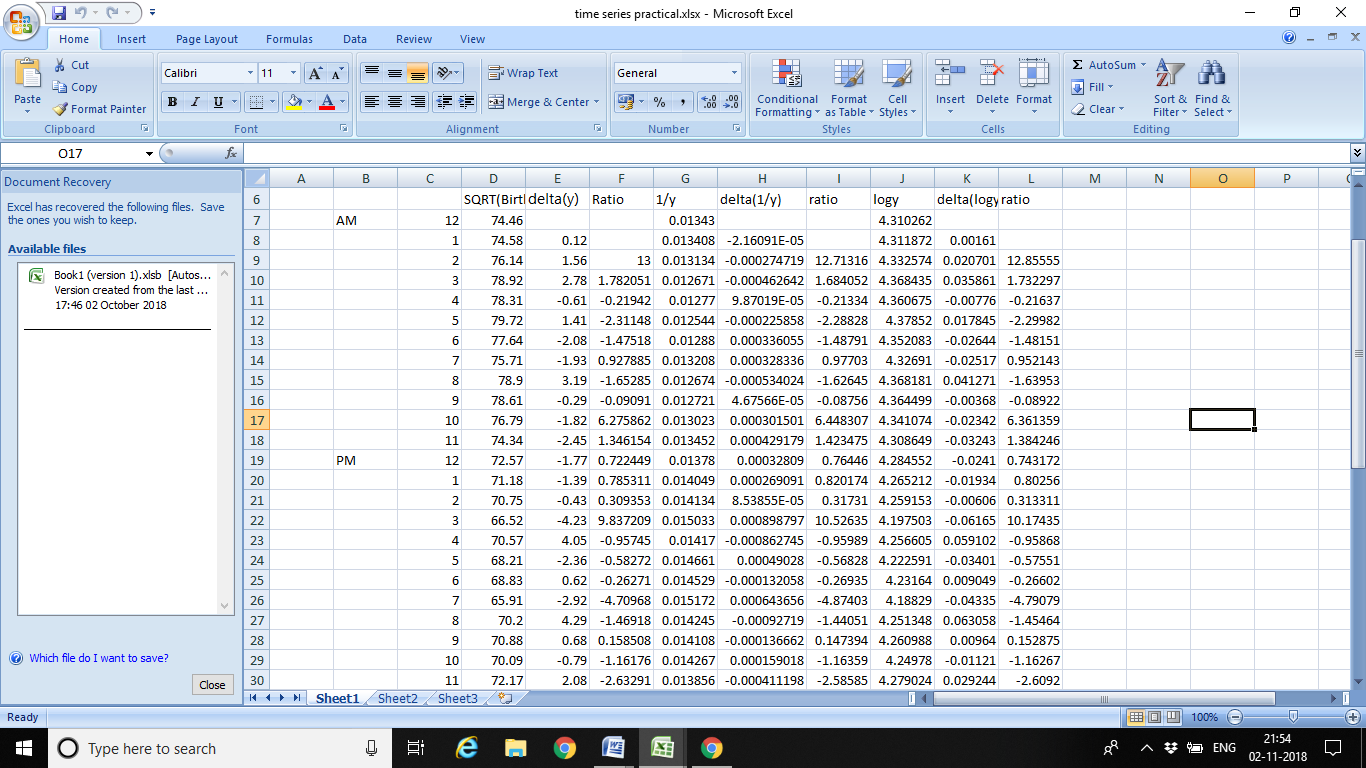
Δy(t)/ Δy(t-h)=c^h , a constant.

Thus the most striking feature of the modified exponential curve is that the first differences of the consecutive values of y(t) corresponding to equivalent values of t change by a constant ratio. This implies that the first differences of y(t) when plotted on a semi logarithmic graph paper, lie on a straight line.

We cannot fit modified exponential curve using principle of least square s. Hence we use method of three selected points or method of partial sums.

Gompertz curve:

If Δ(1/yt)=constant, fit Gompertz curve



LOGISTIC CURVE:

A **logistic function** or **logistic curve** is a common "S" shape ([sigmoid curve](https://en.wikipedia.org/wiki/Sigmoid_function)), with equation:

{\displaystyle f(x)={\frac {L}{1+e^{-k(x-x\_{0})}}}}where

* *e* = the [natural logarithm](https://en.wikipedia.org/wiki/Natural_logarithm) base (also known as [Euler's number](https://en.wikipedia.org/wiki/E_(mathematical_constant))),
* *x*0 = the *x*-value of the sigmoid's midpoint,
* *L* = the curve's maximum value, and
* *k* = the steepness of the curve.[[1]](https://en.wikipedia.org/wiki/Logistic_function#cite_note-1)

If Δlog(yt)=constant, fit a logistic curve.

Seasonal component in the time series:

It is a pattern that reflects regular fluctuations. These short-term movements occur due to the seasonal factors and custom factors of people. In this case, the data faces regular and predictable changes that occurred at regular intervals of calendar. It always consist of fixed and known period.

The main sources of seasonality are given below -

* Climate
* Institutions
* Social habits and practices
* Calendar

How is the seasonal component estimated?

If the deterministic analysis is performed, then the seasonality will remain same for similar interval of time. Therefore, it can easily be modelled by dummy variables. On the other hand, this concept is not fulfilled by stochastic analysis. So, dummy variables are not appropriate because the seasonal component changes throughout the time series.

Different **models** to create a seasonal component in time series are given below -

* **Additive Model -** It is the model in which the seasonal component is added with the trend component.
* **Multiplicative** **Model -** In this model seasonal component is multiplied with the intercept if trend component is not present in the time series. But, if time series have trend component, sum of intercept and trend is multiplied with the seasonal component.

For the data given to us based on no. of births at different hours of the day, it is not possible to calculate seasonal variation.

CYCLIC VARIATIONS IN THE TIME SERIES:

The pattern exhibit up and down movements around a specified trend is known as cyclic pattern. It is a kind of oscillations present in the time series. The duration of cyclic pattern depends upon the industries and business problems to be analysed. This is because the oscillations are dependable upon the business cycle.

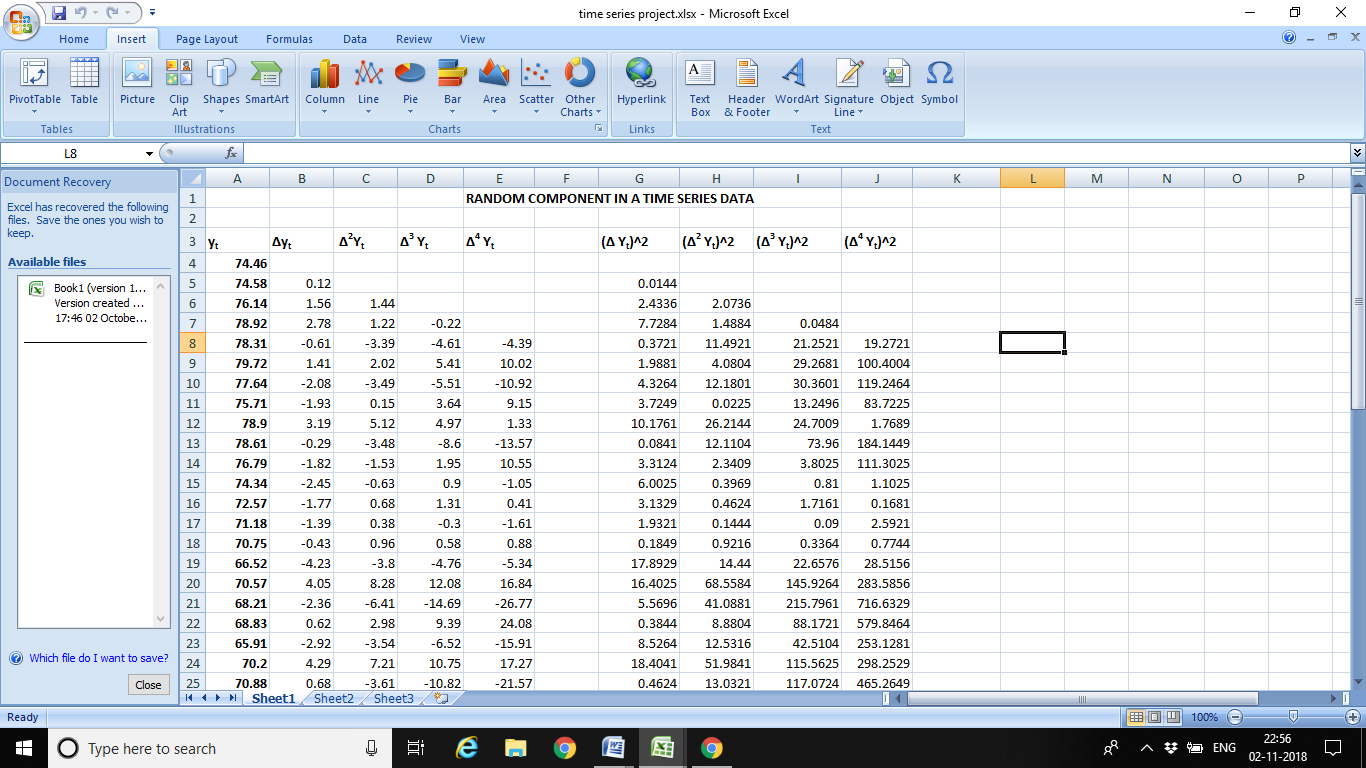
They are larger variations that are repeated in a systematic way over time. The period of time is not fixed and usually composed of at least 2 months in duration. The cyclic pattern is represented by a well-shaped curve and shows contraction and expansion of data.

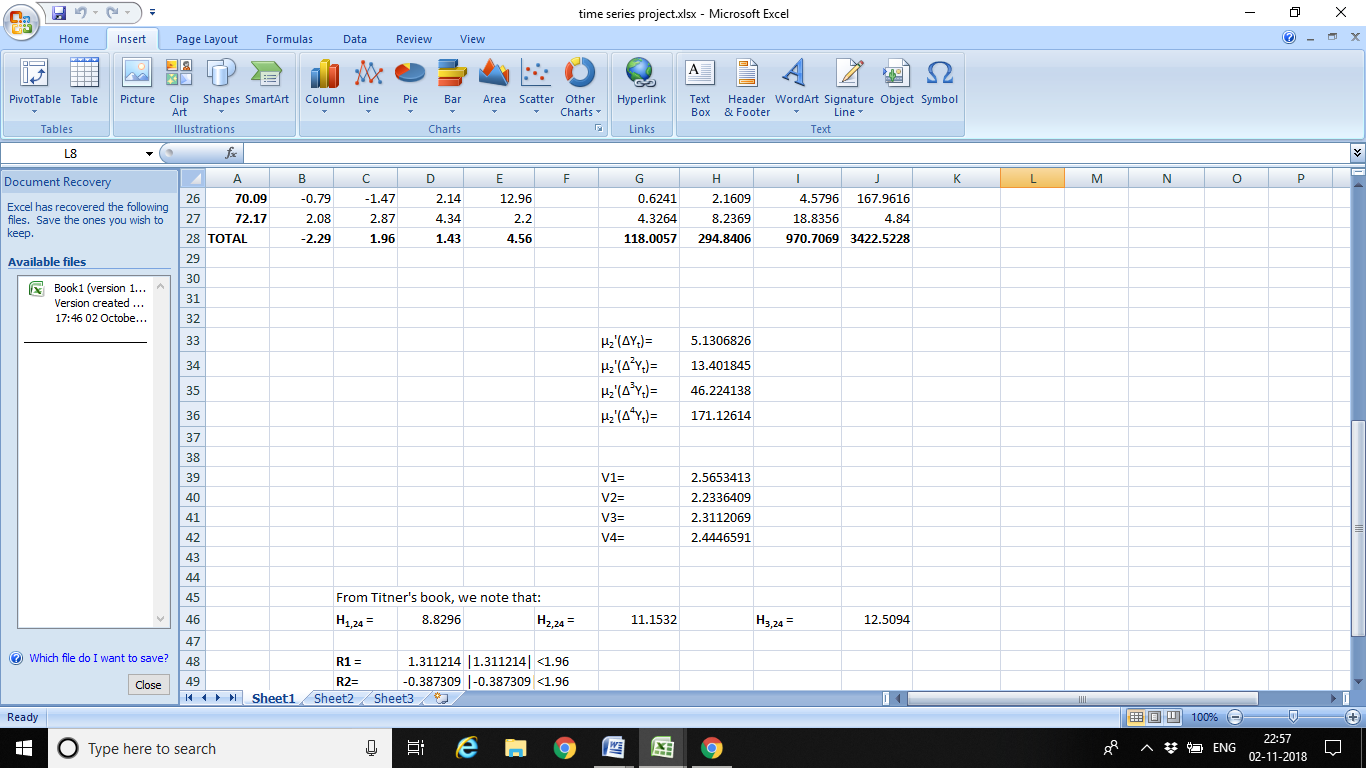
For the data given to us based on no. of births at different hours of the day, it is not possible to calculate cyclic variation.

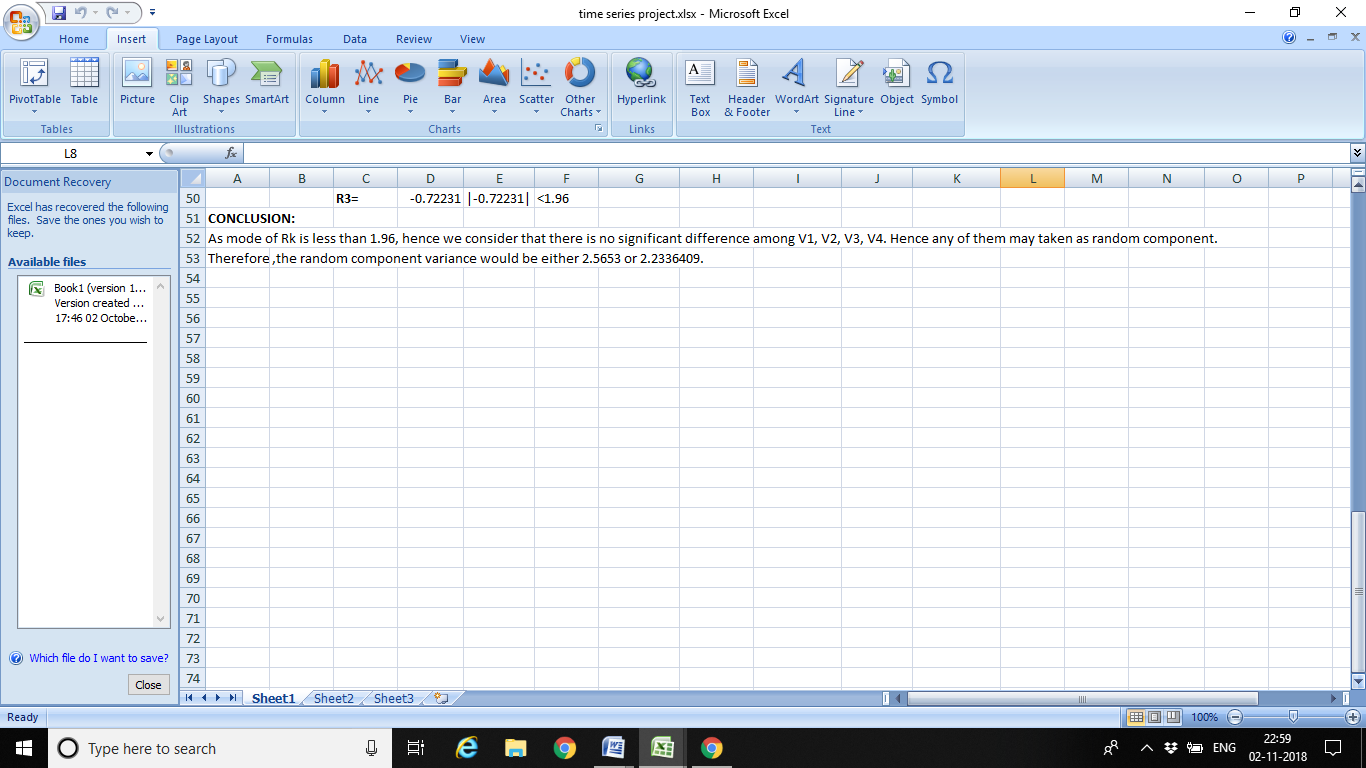
RANDOM COMPONENT OF A TIME SERIES:

It is an unpredictable component of time series. This component cannot be explained by any other component of time series because these variational fluctuations are known as random component. When the trend cycle and seasonal component is removed, it becomes residual time series. These are short term fluctuations that are not systematic in nature and have unclear patterns.

Using the variate difference method, we have calculated the random component of the time series:







Thus, for the given time series data, we have calculated trend using method of moving averages and variance of random component has been calculated using the variate difference method.