Moving Average & Median Smoothing

Smoothing is a technique that involves replacing the original time series data with another one where most of the variation has been removed to see if there is a secular trend. Points are lost at the start and end of the time series.

There are two types of smoothing

1. Moving average (mean) smoothing

Moving-mean smoothing is an option that is preferred for data sets **with few random fluctuations.**

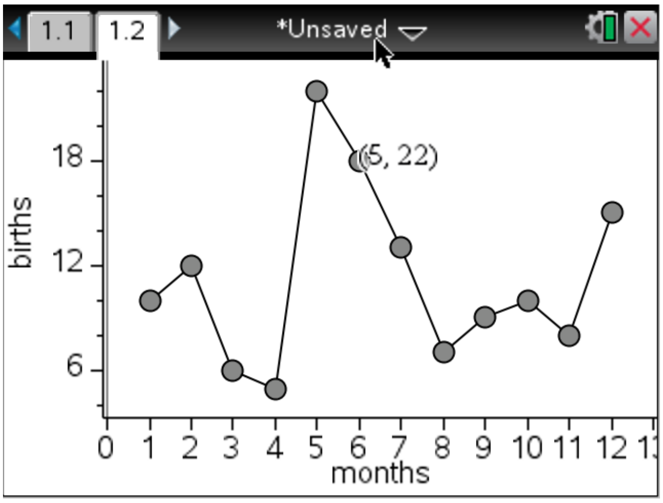
2. Median smoothing

Median smoothing is preferred where there are **small data sets**, as it can be done graphically on a time-series plot. Also, for data sets **with many outliers** due to the **volatile random or cyclical trend**, median smoothing is preferred.

Example

The number of births per month over calendar year in small country hospital is given below:

Month	Number of births
January	10
February	12
March	6
April	5
May	22
June	18
July	13
August	7
September	9
October	10
November	8
December	15

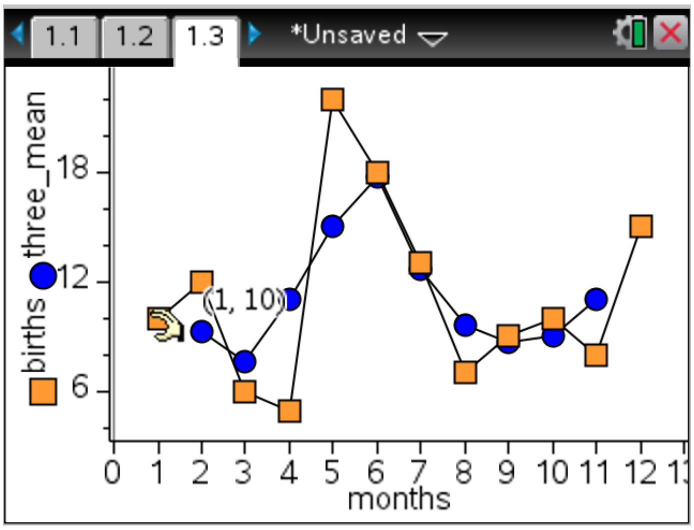


The scatterplot clearly shows considerable variation and fluctuation. Before applying a trend line to this data to make predictions, the data needs to be smoothed.

Attempt.1 - A 3 point moving average smoothing.

Month	Number of births	3 point moving average.
January	10	
February	12	$= \frac{10+12+6}{3} = 9.3$
March	6	$= \frac{12+6+5}{3} = 7.7$
April	5	$= \frac{6+5+22}{3} = 11.0$
May	22	$= \frac{5+22+18}{3} = 15.0$
June	18	$= \frac{22+18+13}{3} = 17.7$
July	13	$= \frac{18+13+7}{3} = 12.7$
August	7	$= \frac{13+7+9}{3} = 9.7$
September	9	$= \frac{7+9+10}{3} = 8.7$
October	10	$= \frac{9+10+8}{3} = 9.0$
November	8	$= \frac{10+8+15}{3} = 11.0$
December	15	

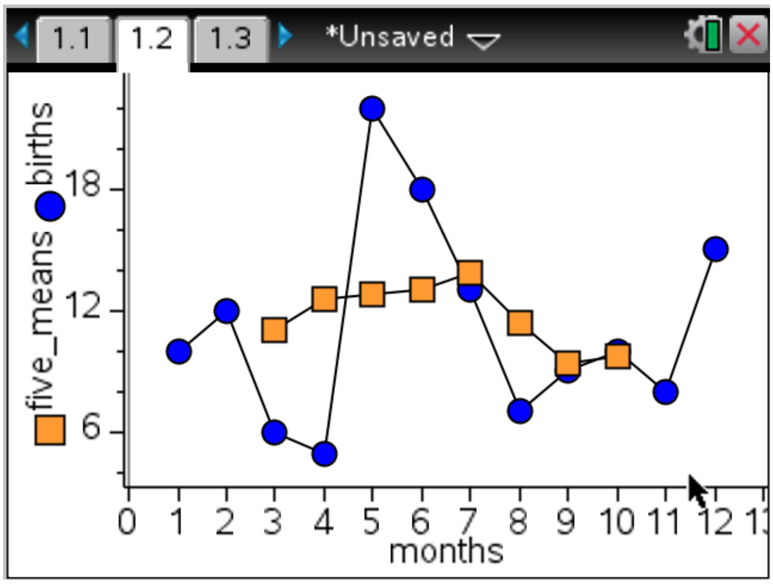
The 3 point moving average smoothing has reduced some variation and fluctuation. However, considerable variation still exists.
NB: A 3 point moving average loses 2 points of data, the first and last month's data.



As can be seen, even following a 3 point moving average smoothing, there still is considerable fluctuation. Another option is to use a 5 point moving average smooth on the original data to better smooth the data.

Attempt.2 - A 5 point moving average smoothing.

Month	Number of births	5 point moving average.
January	10	
February	12	
March	6	$= \frac{10+12+6+5+22}{5} = 11.0$
April	5	$= \frac{12+6+5+22+18}{5} = 12.6$
May	22	$= \frac{6+5+22+18+13}{5} = 12.8$
June	18	$= \frac{5+22+18+13+7}{5} = 13.0$
July	13	$= \frac{22+18+13+7+9}{5} = 13.8$
August	7	$= \frac{18+13+7+9+10}{5} = 11.4$
September	9	$= \frac{13+7+9+10+8}{5} = 9.4$
October	10	$= \frac{7+9+10+8+15}{5} = 9.8$
November	8	
December	15	

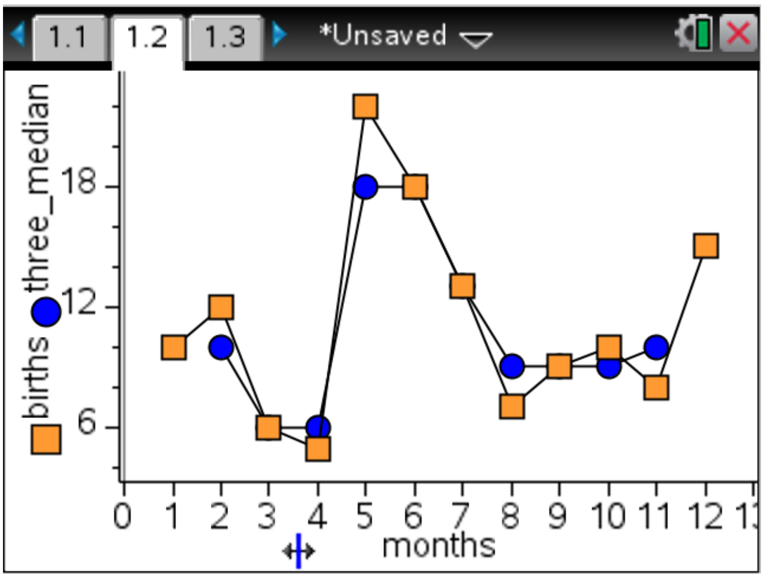


The 5 point moving average smoothing further smoothed the original data. The extreme fluctuations have been removed and a linear equation could be fit, from which predictions can be made.

An alternative to moving average smoothing is **median smoothing**. As the name suggest median smoothing involves finding the median of a set of data. Let's consider the above births per month data and perform a 3 point median smoothing.

Attempt.3 - A 3 point median smoothing.

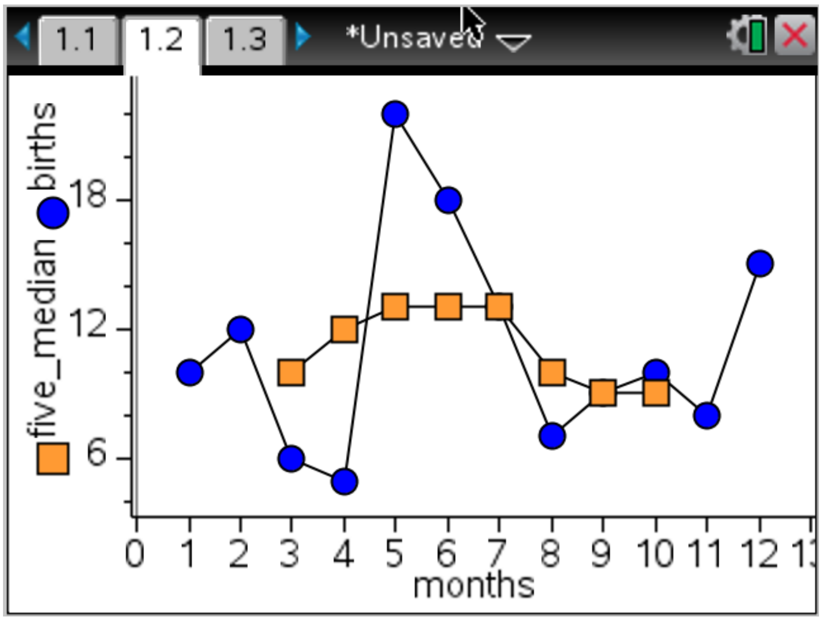
Month	Number of births	3 point median average.
January	10	
February	12	6 **10** 12
March	6	5 **6** 12
April	5	5 **6** 22
May	22	5 **18** 22
June	18	13 **18** 22
July	13	7 **13** 18
August	7	7 **9** 13
September	9	7 **9** 10
October	10	8 **9** 10
November	8	8 **10** 15
December	15	



Following the 3 point median smoothing there has been an obvious reduction in the fluctuation within the data. Once again the number of points plotted have reduced by two as a result of this smoothing process.
Let's now consider a 5 point median smoothing.

Attempt.4 - A 5 point median smoothing.

Month	Number of births	3 point median average.
January	10	
February	12	
March	6	5 6 **10** 12 22
April	5	5 6 **12** 18 22
May	22	5 6 **13** 18 22
June	18	5 7 **13** 18 22
July	13	7 9 **13** 18 22
August	7	7 9 **10** 13 18
September	9	7 8 **9** 10 13
October	10	7 8 **9** 10 15
November	8	
December	15	



As can be seen for the above graph, the 5 point median smoothing technique has removed considerable fluctuation from the original data. Once again in performing a 5 median smooth, 2 pieces of data have been lost from both the first and last month's data (ie. Jan, Feb, Nov & Dec).

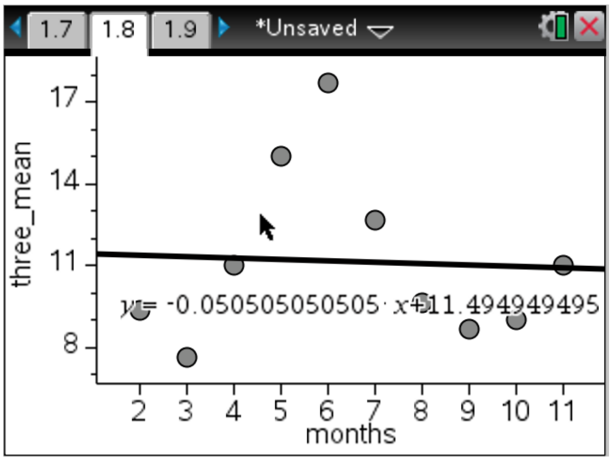
From any one of the four example smoothing techniques shown you can perform a linear regression to create a least squares regression line, from which you can make predictions.

Examples

Use linear regression to predict the number of births in the country hospital for March and November of the following year, using each of the four different methods of smoothing examined.

(**NB:** Round the linear regression equation to 2 decimal places)

1. 3 point moving average smooth

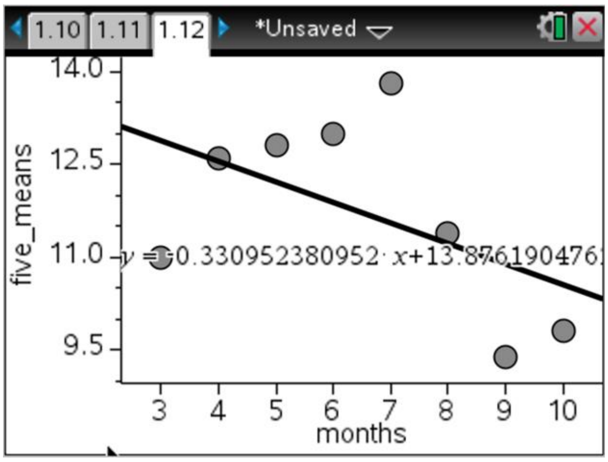


No. Births = -0.05 x Month + 11.49

March of the follow year (Month = 15)
∴ No. Births = -0.05 x (15) + 11.49
= <u>10.74 (approximately 11 births)</u>

November of the follow year (Month = 23)
∴ No. Births = -0.05 x (23) + 11.49
= 10.34(approximately 10 births)

2. 5 point moving average smooth

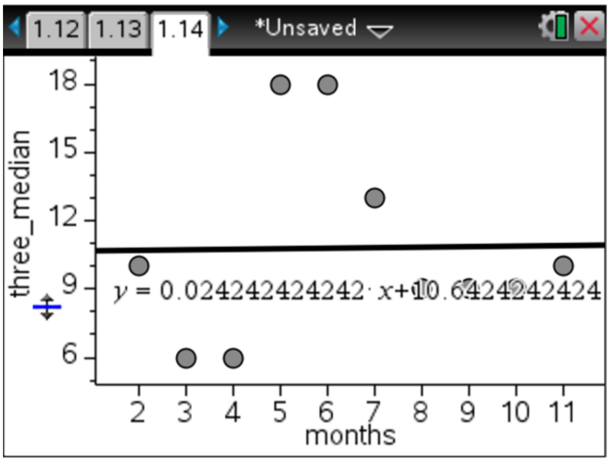


No. Births = -0.33 x Month + 13.88

March of the follow year (Month = 15)
∴ No. Births = -0.33 x (15) + 13.88
= <u>8.93 (approximately 9 births)</u>

November of the follow year (Month = 23)
∴ No. Births = -0.33 x (23) + 13.88
= <u>6.29 (approximately 6 births)</u>

3. 3 point median smooth

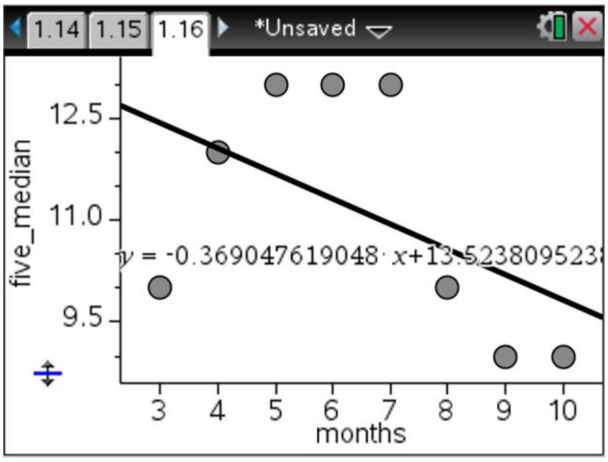


No. Births = 0.02 x Month + 10.64

March of the follow year (Month = 15)
∴ No. Births = 0.02 x (15) + 10.64
= 10.94 (approximately 11 births)

November of the follow year (Month = 23)
∴ No. Births = 0.02 x (23) + 10.64
= 11.1 (approximately 11 births)

4. 5 point median smooth



No. Births = -0.37 x Month + 13.52

March of the follow year (Month = 15)
∴ No. Births = -0.37 x (15) + 13.52
= 7.97 (approximately 8 births)

November of the follow year (Month = 23)
∴ No. Births = -0.37 x (23) + 13.52
= 5.01 (approximately 5 births)