

Mean & Median Smoothing

Teacher Notes & Answers

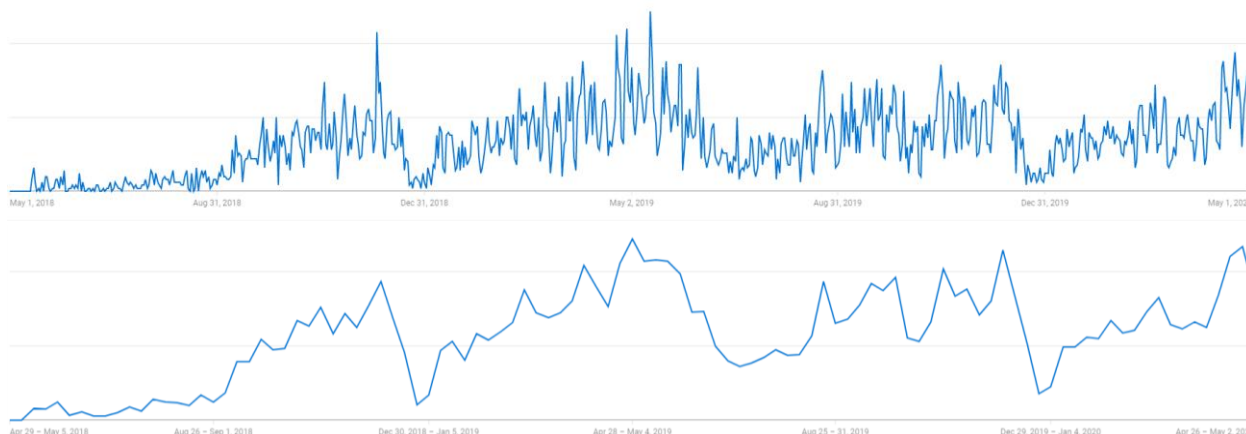
7 8 9 10 11 12



Introduction – From YouTube to Photoshop

The charts below show the number of times a video on TI-Nspire was viewed on YouTube over a period of two years since the video was originally uploaded (May 2018). The top graph shows daily views and displays a lot of ‘noise’ in the form of daily fluctuations. To help identify general trends the second graph shows weekly views. The weekly view is a method of smoothing the data, effectively averaging out the views over a 7 day period.

Both charts show that the number of views is lower from December through to January; the second chart really shows how rapid the decline is from the start of December. As the content of the video is related to high school mathematics, the drop over the December - January period is not entirely unexpected. It is interesting to note that the weekly data seems to display a significant point in time when the views decline. This decline aligns remarkably well to the end of Year 11 and 12 studies each year. What other dips are evident? What is your forecast for June – July 2020?



Teacher Notes:

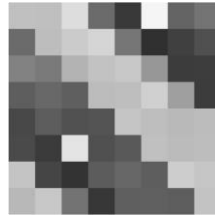
There are many data sources that have the ability to change between daily / weekly / monthly time periods that in turn generate a level of smoothing. While the statistical means used to smooth is different, the purpose is the same. The ‘daily’ data (above) shows two specific spikes:

- November 2018
- May 2019.

The November spike occurred immediately prior to end of year state-wide examinations. So, what cause the spike in May? An IB exam was scheduled during this time. So the daily data can provide some interesting individual elements whilst the ‘smoothed’ data helps show the general trends. What remains is the statistical “means” by which the smoothing is obtained.

Rather than focusing on traditional data, it is important for student to realise that other processes rely on smoothing approaches. The example selected here is ‘image enhancement’. By looking at the visual differences between mean and median smoothing students can see why alternatives are necessary. The visual differences encourage students to consider how this result is being achieved in addition to helping them appreciate statistical understanding and knowledge.

Median smoothing is a method used by photo-imaging suites to automatically touch up a picture. The simplest way to understand this process is to consider a grey-scale image. Every pixel is represented by a single number between 0 (black) and 255 (white). The 300 x 400 pixel image below (left) has deteriorated with age. The image in the middle is an enlarged portion of the top left corner of the deteriorated image; immediately below are the numbers that generate this section. A pixel in the top row is almost white (245). To the left and right are the numbers 57 and 100. Below 245 are the numbers 118, 48 and 64. These numbers are deemed as being “immediately adjacent”. The median of this data set: {48, 57, 64, 100, 118, 245} is $(64 + 100) \div 2 = 82$. The almost white pixel (245) will be replaced with dark grey (82).



192	191	221	107	57	245	100	116
107	189	203	213	118	48	64	80
130	122	179	194	200	151	60	61
94	104	95	188	190	208	159	60
72	93	81	84	197	190	185	188
74	59	228	80	89	192	187	189
196	69	51	92	103	94	205	198
184	200	115	55	95	96	94	195



Computer imaging software can scan every pixel and change each one based on the numbers immediately surrounding the pixel. The result is to correct all the drop outs in the deteriorated image with a view to returning the photo to its original glory, the photo shown on the right (previous page).

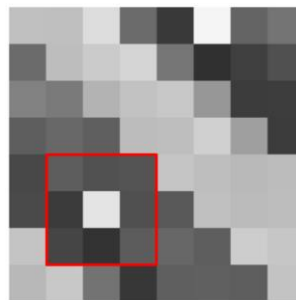
Teacher Notes:

The purpose of these two questions is to show students that a numerical, statistical approach can be used to automate the process associated with image enhancement. In both the median and mean cases the replaced pixel more closely resembles the surrounding pixels. Students should note that in Question 2 the pixel is at the upper end of the surrounding pixels due to the mean being impacted by the outlier. (228)

Question: 1.

Another pixel in the enlarged section is close to white (228) whilst the eight surrounding pixels are relatively dark: {93, 81, 84, 59, 69, 51, 92} (see below). Determine the median value for the nine pixels and hence the new value for the 228 pixel.

192	191	221	107	57	245	100	116
107	189	203	213	118	48	64	80
130	122	179	194	200	151	60	61
94	104	95	188	190	208	159	60
72	93	81	84	197	190	185	188
74	59	228	80	89	192	187	189
196	69	51	92	103	94	205	198
184	200	115	55	95	96	94	195



Answer:

Arranged in ascending order: 51, 59, 69, 80, 81, 84, 92, 93, 228. Median: 81
New pixel value will therefore be 81.

Question: 2.

Using the same set of pixels as the previous question, determine the replacement value for the 228 pixel if ‘mean’ smoothing was applied rather than median.

Answer:

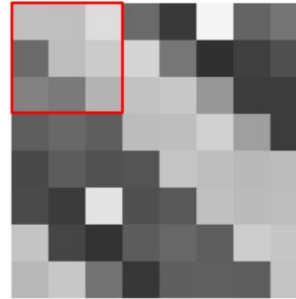
No requirement to order the values. Mean: 93
New pixel value will therefore be 93.

In the absence of filtering techniques that identify 'bad' pixels, every pixel would naturally undergo median or mean filtering, this has implications for the remaining image. This section of the image has a distinctive light and dark band. The following questions explore what happens to the image where the light band transitions to the dark band and vice-versa.

Question: 3.

Determine the result of median and mean smoothing on the centre (189) pixel and comment on the result.

192	191	221	107	57	245	100	116
107	189	203	213	118	48	64	80
130	122	179	194	200	151	60	61
94	104	95	188	190	208	159	60
72	93	81	84	197	190	185	188
74	59	228	80	89	192	187	189
196	69	51	92	103	94	205	198
184	200	115	55	95	96	94	195



Answer:

Arranged in ascending order: 107, 122, 130, 179, 189, 191, 192, 203, 221.

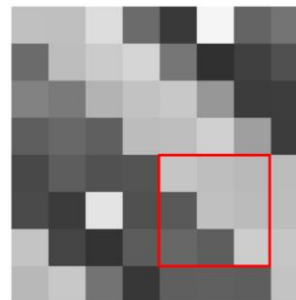
Median: 189 Mean: 170

Despite the fact that there are no outliers amongst the data there is a relatively large difference between the mean and median values. It is interesting to note that median smoothing will not change the value of the centre pixel. Mean smoothing however will make the pixel a little darker.

Question: 4.

Determine the result of median and mean smoothing on the centre (192) pixel and comment on the result.

192	191	221	107	57	245	100	116
107	189	203	213	118	48	64	80
130	122	179	194	200	151	60	61
94	104	95	188	190	208	159	60
72	93	81	84	197	190	185	188
74	59	228	80	89	192	187	189
196	69	51	92	103	94	205	198
184	200	115	55	95	96	94	195



Answer:

Arranged in ascending order: 89, 94, 103, 185, 187, 190, 192, 197, 205.

Median: 187 Mean: 160

Just like Question 3, the mean smoothing makes the biggest change to the centre pixel. The pixel in Question 3 and also the one in Question 4 lie on what appears to be a border of significant contrast. In both cases mean smoothing has made the largest adjustment effectively 'blurring' the border.

Automating the process

Open the TI-Nspire file: Image Enhancement

Page 1.1 contains a Notes Application providing some detail about the size of the image that has been embedded in this file. The image is a small section of the image shown on page 1 of this worksheet.

Read through the instructions on Page 1.1 and then navigate to Page 1.2.

Select the drawing program from the VAR menu:

```
drawing(100)
```

When the drawing program is launched from Page 1.2, two versions of the image will be displayed on the drawing canvas.

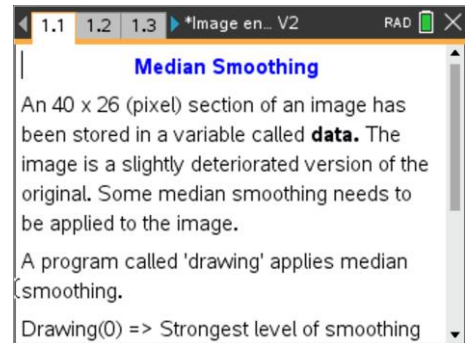
The image on the left is part of the original. The image on the right has been filtered, however by using: drawing(100), no filtering has actually been completed.

Can you identify which region of the original image has been captured?

To return to the calculator application where the program was launched, press [esc].

It is now time to allow median smoothing to repair the pixels. Run the Drawing program again as:

```
Drawing(0)
```



Question: 5.

Did the filtering tool remove all the drop-out pixels? What else changed about the image? Explain.

Answer:

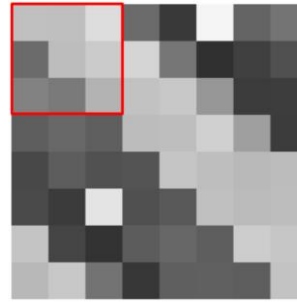
All the drop out pixels are removed. Each drop out pixel is significantly different to the surrounding area so it is replaced with the median value. The problem is that the replacement is indiscriminant which means overall there is less variation from pixel to pixel, from a visual perspective this means the contrast is lost and the image appears blurred. Detail is lost from some areas, for example the fascia board on the verandah had a light horizontal band correlating to where sunlight may have been striking it, this band is completely erased.

It would be nice if the data smoothing only applied to points that were clearly indicative of deterioration or pixel drop-out. The number used in the 'drawing' program provides such an opportunity. The program continues to run median smoothing, however the number determines just when it will happen. The number determines how many 'standard deviations' the current pixel is away from the mean of the surrounding pixels.

Question: 6.

To help understand how this parameter affects this selective filtering process we return to the set of pixels investigated in Question 3.

192	191	221	107	57	245	100	116
107	189	203	213	118	48	64	80
130	122	179	194	200	151	60	61
94	104	95	188	190	208	159	60
72	93	81	84	197	190	185	188
74	59	228	80	89	192	187	189
196	69	51	92	103	94	205	198
184	200	115	55	95	96	94	195



- a) Determine the mean and standard deviation of this set of numbers corresponding to the cluster of 9 pixels.

Answer: Data: 192,191,221,107,189,203,130,122,179

Mean: 170.4

Standard Deviation: 40.2 (Sample)

- b) If drawing(10) is used, then the middle pixel remains unchanged if it falls within the following range:

$$\bar{x} - \sigma \leq x \leq \bar{x} + \sigma$$

Use your answers to part (a) to determine whether or not 189 is outside this range and therefore the value of this pixel once this form of selective filtering is completed.

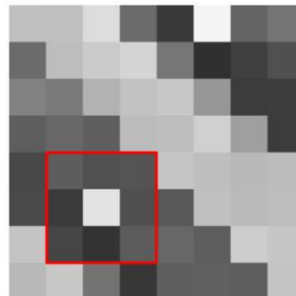
Answer: $170.4 - 40.2 \leq x \leq 170.4 + 40.2 \dots 130.2 \leq x \leq 210.6$

The 189 pixel lies within one standard deviation of the mean and would therefore NOT be changed if mean filtering with 'one standard deviation' was implemented.

Question: 7.

The region studied in Question 1 contains a pixel that dropped out.

192	191	221	107	57	245	100	116
107	189	203	213	118	48	64	80
130	122	179	194	200	151	60	61
94	104	95	188	190	208	159	60
72	93	81	84	197	190	185	188
74	59	228	80	89	192	187	189
196	69	51	92	103	94	205	198
184	200	115	55	95	96	94	195



- a) Determine the mean and standard deviation of this set of numbers corresponding to the cluster of 9 pixels.

Answer:

Mean: 93.

Standard Deviation: 52.62

- b) If drawing(10) is used, then the middle pixel remains unchanged if it falls within the following range:

$$\bar{x} - \sigma \leq x \leq \bar{x} + \sigma$$

Use your answers to part (a) to determine whether or not 228 is outside this range and therefore the value of this pixel once this form of selective filtering is completed.

Answer: $93 - 52.62 \leq x \leq 93 + 52.62 \dots 40.4 \leq x \leq 145.6$

The 228 pixel does not lie within one standard deviation of the mean and would therefore still be changed if mean filtering with 'one standard deviation' was implemented.

Question: 8.

Explain why the standard deviation might be used as a guiding statistic rather than Interquartile range.

Answer: The relatively small data set means that 'quartiles' will consist of essentially 2 elements. The small sample size can also generate significant bias. For example, the data set from the previous imaging question:

Median: 81 { 51, 59, 69, 80, 81, 84, 92, 93, 228 }

Q₁: 69 { 51, 59, 69, 80, 81 }

Q₃: 92 { 81, 84, 92, 93, 228 }

Note that the number: 81 is included in the calculation of Q₁ and Q₃. This creates a bias in the results. The interquartile range is equal to $92 - 69 = 23$.

Question: 9.

Explain why the deviation from the 'mean' rather than the 'median' might be used in this refined process.

The standard deviation measure 'deviations' from the mean, therefore it is more appropriate to use the mean rather than the median in association with computations pertaining involving the standard deviation.

Teacher Notes:

The Median Absolute Deviation (MAD) could be used rather than the standard deviation in each of the calculations; however MAD computations are not included in most courses. The focus of the question is therefore to draw student attention to the fact that the standard deviation measures variability about the 'mean'.

Question: 10.

Run the filtering tool again using: drawing(10). Compare the original image (with pixel drop-outs) to the filtered image. Did the filtering still remove the drop-outs? (Explain) What other differences can you see when this added restriction is placed on the filtering.

All drop-out pixels removed with a level 10 filtering applied. The replacement pixels blend perfectly with their surrounding environment. The smoothed image is not as sharp as the original image; this is because many of the pixels along the 'edge' of an object are adjusted. (As per Question 5)

**Teacher Notes:**

Level 10 filtering means that any points that lie beyond 1 standard deviation from the mean will be adjusted.

Question: 11.

Run the filtering tool again using: drawing(20). Once again compare the original image (with pixel drop-outs) to the filtered image. What differences do you notice this time?

Once again the drop-out pixels are removed since they are at least two standard deviations from the mean. Other less substantial pixels however remain unchanged therefore leaving a reasonable detail and contrast in the enhanced image.



Further Deterioration

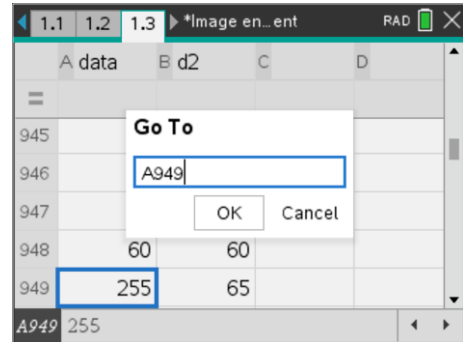
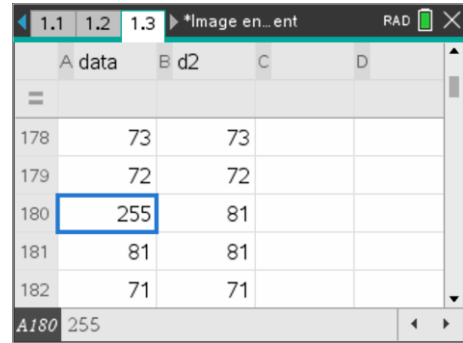
What happens if our image has deteriorated further? If there are more pixel dropouts 'clusters' of dropouts become increasingly more likely. The existing dropouts are in the following locations:

- Pixel: 180
- Pixel: 495
- Pixel: 910

Adding dropouts adjacent to these existing locations makes it possible to see how clustering affects the filtering process.

The following pixels changes need to be made to the image 'data'.

- Pixel: 181 Change to: 255
- Pixel: 535 255
- Pixel: 949 255



To navigate to a specific cell in the spreadsheet: Press Ctrl + G and enter the cell reference. For example entering: **A949** would set the current cell to one of the pixels that needs to be altered.

Question: 12.

With the increased 'damage' to the original image, run the filtering tool again using: drawing(20).

- a) Did all the pixel drop-outs get fixed? Explain.

None of the pixels are corrected with this level of filtering. Now that there are two points of significant difference in the sample, the standard deviation would be larger therefore the limitations imposed by the filter are not significant enough to identify these 'double' pixel drop-outs as worthy of replacement.

- b) Try running the program again as: drawing(10). Did all the pixels get fixed, if so, at what 'cost'?

All the pixels were corrected, although they do leave a noticeable 'foot-print' or residue. The first pair has a light grey replacement, similarly with the second pair, however the third pair is almost completely untraceable.

The cost is a loss in contrast as more pixels fall under the restrictions.



- c) Try using other drawing values such as 17 or 15. What value removes the dropout pixels with minimal compromise to the quality of the remaining image?

Removal by different levels of filtering (shown below) illustrates that the corrected pixels are still noticeable and that level 15 filtering is the first to completely remove all drop-outs. This level of filtering also retains a reasonable level of contrast. This experimentation explains why users of imaging software need to experiment with the filtering tools for each and every image so as to determine the ideal level applicable to their image.



Drawing(15)



Drawing(16)



Drawing(17)

Question: 13.

Suppose a new median filtering tool was created that sampled an additional layer of numbers around any given pixel. (Sample of 25). Discuss how this might affect filtering and image quality.

The larger sample size would work well on higher resolution images, however on lower resolution or pixel concentration, the larger sample is more likely to include regions of significant contrast. Pixel drop-outs close to regions of high contrast will be more difficult to remove. Even the sampling from Question 12 illustrates that the 'hardest' pixel to remove [Drawing(16)] is located in an area where an edge generates greater pixel value variation.

Question: 14.

Design a new statistically based filtering tool that might help eliminate larger 'damaged' regions of an image without losing other aspects of the image quality.

Answers will vary.

Teacher Notes:

The purpose of Question 14 is for students to realise they need to *understand* a large range of statistical methods in order to be able to design a filtering process. The question also provides an opportunity for students to articulate their understanding. Students may choose to tackle the situation highlighted by Question 12 where 'double drop-outs' occur. The standard deviation increases as a result of the two drop-outs in one region. A larger sample could provide for the use of IQR measurements. Students could also discuss a method of counting the number of pixels outside a specific range and identifying how far they are from the rest of the pixels.

Example:

Data Set 1: {14, 14, 14, 14, 14, 14, 14, 14, 24, 24} Mean: 16 St.Dev: 4.22

Data Set 2: {7, 9, 12, 13, 14, 16, 18, 21, 22, 28} Mean: 16 St.Dev: 6.39

Data set 1 has two values that may be 'drop-outs' but would not be picked up using standard deviations.