DATA ANALYSIS -- THE SUM OF A SERIES

Consider the series: $1^2 + 2^2 + 3^2 + ... + n^2$. Is there a relationship between the number of terms being added (n), and the sum of those terms?

1. To investigate, let's make a table of the sum of 1 term, 2 terms, 3 terms, ... 8 terms, to see if we can recognize a pattern:

# of terms	sum of n terms
1	
2	
3	
4	
5	
6	
7	
8	

2. Clear lists L_1, L_2 and enter the number of terms into L_1, and the sum of n terms into L_2. Use an appropriate window and graph this data.

3. Discuss what type of relationship, what type of curve, this could be.

4. Let's try a quadratic regression. Have the calculator calculate a quadratic regression of the data in L_1 and L_2 and store the regression equation into y_1. Graph it to see how good a "fit" it is.

5. The real test is how close the function approximates the actual sum of the series. Using y_1, find the approximate sum for:

a) 3 terms b) 5 terms c) 8 terms

d) How good of an approximation is this?

6. We could continue by trial and error to find all the different regression equations until we find the "best fitting" one. But that can be tedious. It would be nice if we could come up with a procedure that would give us a hint as to which regression equation to try first. There is such a procedure that has been around for many years but it was first shown to me by a math teacher from Scotland, Jim Reid. The procedure is explained below.
First, take the table you created originally and add about 4 columns on the right of it, as shown on the next page:

6.

# of terms	sum of n terms	1st difference	2nd difference	3rd difference	4th difference
1	1				
2	5				
3	14				
4	30				
5	55				
6	91				
7	140				
8	204				

In the column called "1st difference", place the difference between successive terms of the previous (sum of n terms) column. For example, 5 - 1 =4; 14 - 5 = 9; ... Continue for the rest of the column.

In the column called "2nd difference", place the difference between successive terms of the previous ("1st difference") column.

Continue to proceed through each column in this manner until each difference in that column is the same constant. That is when you stop.

If the "1st difference" column is the column that has all the same constant, that means that a linear polynomial (or 1st power) equation best "models" this behavior.

If the "2nd difference" column is the column that has all the same constant, that means that a quadratic polynomial (or 2nd power) equation best "models" this behavior.

And so on...

Since this procedure was first told to me by Jim Reid from Scotland, I have nicknamed this "Jim's Maneuver."

Perform "Jim's Maneuver" and see what results you get.

7. Let's find this cubic polynomial by finding the values for A, B, C, D in the general cubic equation: $y = Ax^3 + Bx^2 + Cx + D$. We will do this by substituting 4 of the original ordered pairs in for x and y and generate a system of 4 equations in the 4 variables A, B, C, D. We could use any 4 points -- it is your choice. Find the system of equations now.

Then solve this system of equations by any appropriate method. Convert your values for A, B, C, D into fractions.

8. Turn off y_1, and enter this new equation into y_2. Graph it to see how well it approximates the data.

The real test, though, is how well does it approximate the data. Test it now.

9. Take the expression you have in y_2, and combine that into a single fraction and then factor the numerator.

10. Just out of curiosity let's have the calculator generate the cubic regression to see how closely it matches our equation. Do it. Store the result in y_3.

11. We can now use this equation (formula) to find the sum of any number of perfect squares:

a) Find the sum of the first 20 perfect squares.
b) Find the sum of the first 50 perfect squares.
c) Find the sum of the first 100 perfect squares.
d) Find the sum of the 21st through the 50th perfect squares.

DATA ANALYSIS -- THE SUM OF A SERIES
Teacher Notes and Ideas

1.

# of terms	sum of n terms
1	1
2	5
3	14
4	30
5	55
6	91
7	140
8	204

2.

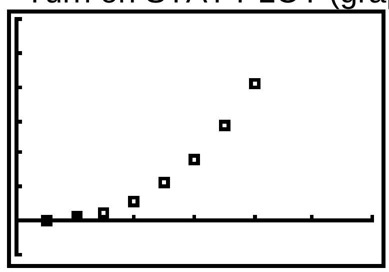
Enter data in L_1, L_2

```
L1     L2     L3    10
3      14
4      30
5      55
6      91
7      140
8      204
------
L2(9) =
```

Set an appropriate window

```
WINDOW
 Xmin=0
 Xmax=12
 Xscl=2
 Ymin=-50
 Ymax=300
 Yscl=50
 Xres=1
```

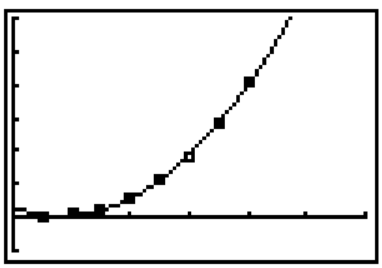
Turn on STAT PLOT (graph)



3. It could be a parabola, a cubic, a quartic, an exponential, ...

4.

```
QuadReg
 y=ax²+bx+c
 a=5
 b=-17
 c=16.5
 R²=.9982255202
```



This looks pretty good. R^2 is very close to one, so this should approximate the sum of n terms fairly well.

5.

```
Y1(3)
                10.5
Y1(5)
                56.5
Y1(8)
               200.5
```

For 3 terms we should get 14, not 10.5. For 5 terms we should get 55, not 56.5. And for 8 terms we should get 204, not 200.5.

These are decent approximations but not exact.

6.

# of terms	sum of n terms	1st difference	2nd difference	3rd difference	4^{th} difference
1	1	4	5	2	
2	5	9	7	2	
3	14	16	9	2	
4	30	25	11	2	
5	55	36	13	2	
6	91	49	15		
7	140	64			
8	204				

***3rd difference is the charm

Therefore, according to Jim's Maneuver, a cubic polynomial will best approximate this data.

7. I will use the first 4 points:

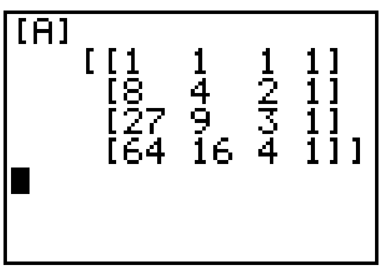
(1,1) $A(1)^3 + B(1)^2 + C(1) + D = 1$

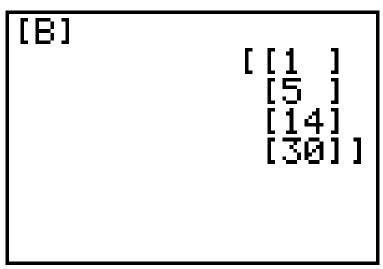
(2,5) $A(2)^3 + B(2)^2 + C(2) + D = 5$

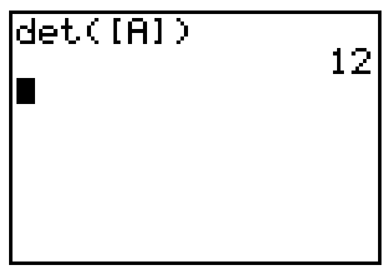
(3,14) $A(3)^3 + B(3)^2 + C(3) + D = 14$

(4,30) $A(4)^3 + B(4)^2 + C(4) + D = 30$

We can now solve this system of 4 equations in 4 variables using the "inverse of a matrix" method:



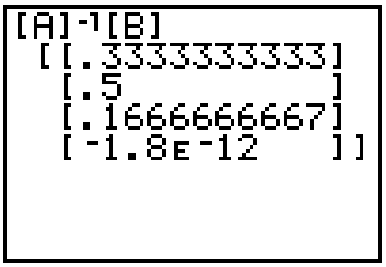


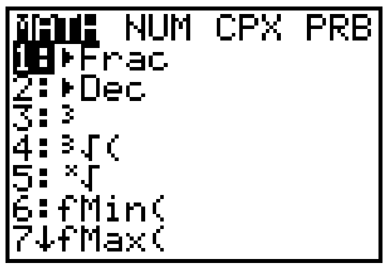


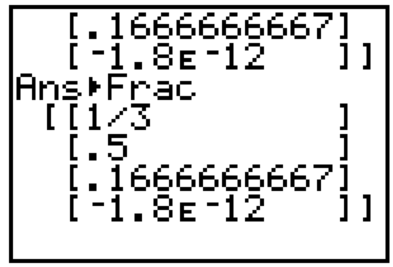
Since the determinant of matrix [A] is non-zero, that means that the inverse of matrix [A] exists and that the system can be solved by the inverse of a matrix method. See the next page for the "theory" that students should remember.

Have students recall:

$$AX = B$$
$$A^{-1} \cdot AX = A^{-1} \cdot B$$
$$I \cdot X = A^{-1} \cdot B$$
$$X = A^{-1} \cdot B$$

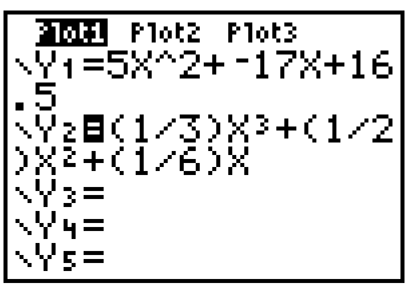


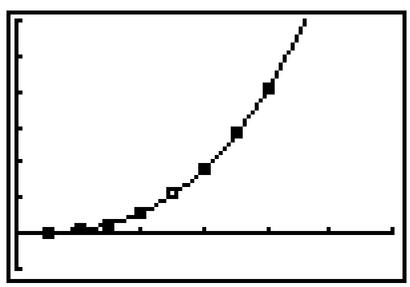




Notice that $A = \frac{1}{3}$ $B = \frac{1}{2}$ $C = \frac{1}{6}$ $D = 0$. Discuss how .16666666... is really $\frac{1}{6}$ and you should have already discussed with your students that -1.8 X 10^{-12} is considered zero for practical purposes.

8. Therefore our equation is $y = \frac{1}{3}x^3 + \frac{1}{2}x^2 + \frac{1}{6}x$. Type this into y_2, turn off y_1 and graph y_2 to see how well it approximates the data.





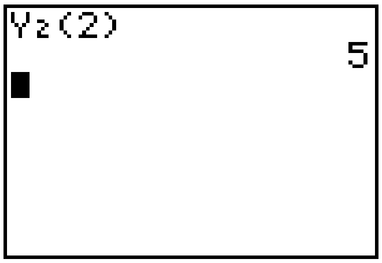
It seems to "fit" the data pretty well.

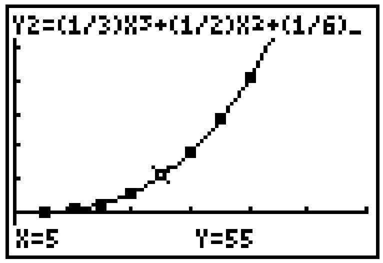
Evaluate y_2 at x = 2, 5, 8, ... to see how well it models the data. You can do this several ways:

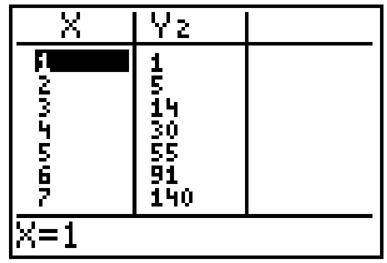
1) By evaluating y_2 at x = 2

2) By tracing on the graph of y_2

3) By looking in a table with x and y_2







This polynomial models this data *EXACTLY*! Oh wow!

9. $$\frac{1}{3}x^3 + \frac{1}{2}x^2 + \frac{1}{6}x = \frac{2}{6}x^3 + \frac{3}{6}x^2 + \frac{1}{6}x$$

$$= \frac{x}{6}(2x^2 + 3x + 1)$$

$$= \frac{x}{6}(2x+1)(x+1)$$

$$= \frac{x(x+1)(2x+1)}{6}$$ *This is how it is normally written.*

Point out to your students that this formula can be found in any Calculus text in the chapter discussing Riemann sums.

I prefer to have my students now write the formula using function notation as follows:

$$S(n) = \frac{n(n+1)(2n+1)}{6} \quad \forall n \in N$$

If you are really feeling ambitious, you can have your students prove this is true using the principle of math induction.

10.

```
CubicReg
 y=ax³+bx²+cx+d
 a=.3333333333
 b=.5
 c=.1666666667
 d=0
 R²=1
```

Oh wow again!

11.

a, b, c)

```
Y3(20)
                2870
Y3(50)
               42925
Y3(100)
              338350
```

d)

```
Y3(50)-Y3(20)
               40055
```