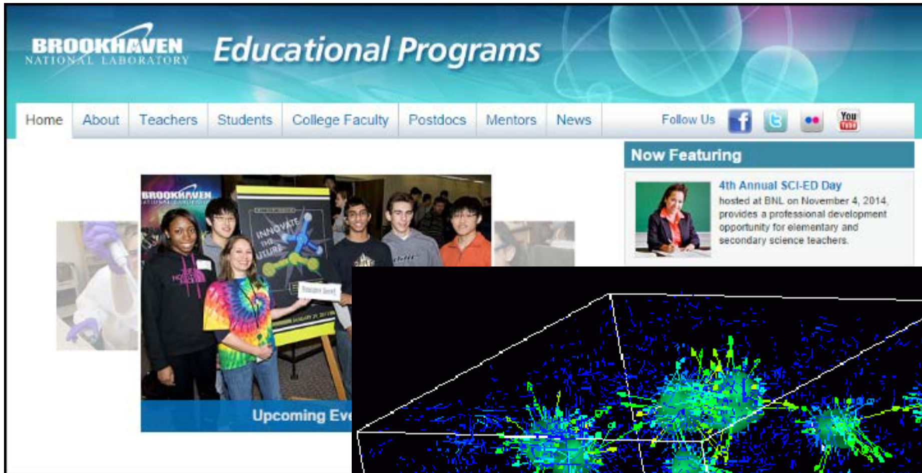
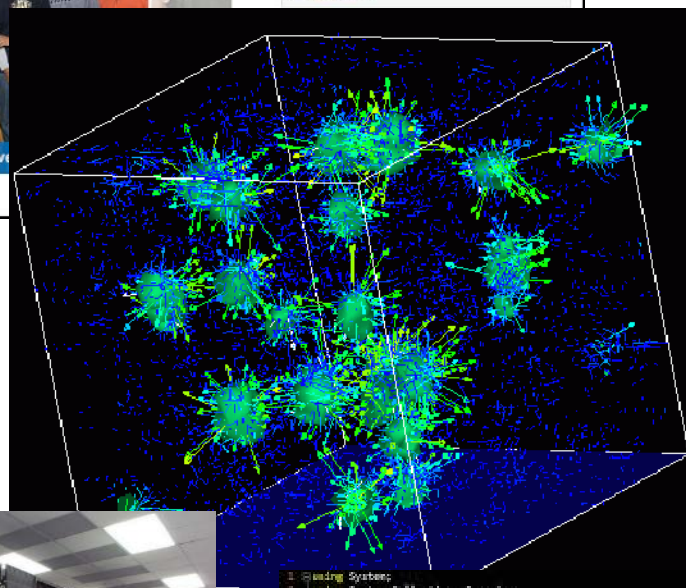
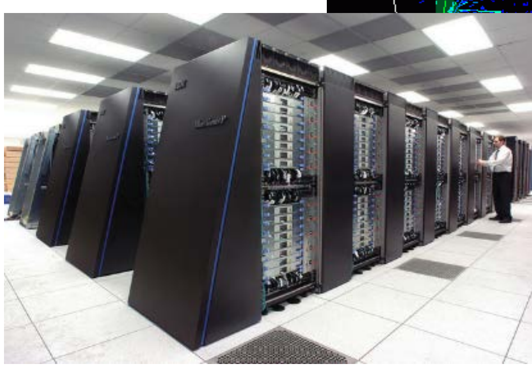






Survey of Scientific Computing

(SciComp 301)

Dave Biersach
Brookhaven National Laboratory
dbiersach@bnl.gov

Session 23

Difference Tables,
Least Squares

Session Goals

- Determine the underlying **generator** for a sequence of integers (the ***functional equation***)
    - Fit a **quadratic** curve to a set of observations to **interpolate** the resulting values that lie *between* those observations
    - Understand how to create **difference tables** in the open source (and free) LibreOffice **Calc** spreadsheet program
    - Appreciate **% relative error** as a measure of the **goodness of fit** between a model and the experimental data
- Fit a curve using the **Method of Least Squares**
    - Derive the least squares equations using **partial derivatives** to calculate the coefficients of a quadratic model

Difference Tables – The Big Picture

- A difference table calculates the delta between successive values (in the **range**) of a given function
- For every higher power of the *independent* variable (in the **domain**) we add another difference column
- Difference column #2 is the gap between successive values of difference column #1, etc.
- We keep adding difference columns until the value in the **rightmost** column is the same for every row – this is called achieving a **steady state**

Difference Tables – The Big Picture

44						
45	Quadratic	y=x^4				
46	x	y	Diff 1	Diff 2	Diff 3	Diff 4
47	1	1				
48	2	16	15			
49	3	81	65	50		
50	4	256	175	110	60	
51	5	625	369	194	84	24
52	6	1296	671	302	108	24
53	7	2401	1105	434	132	24

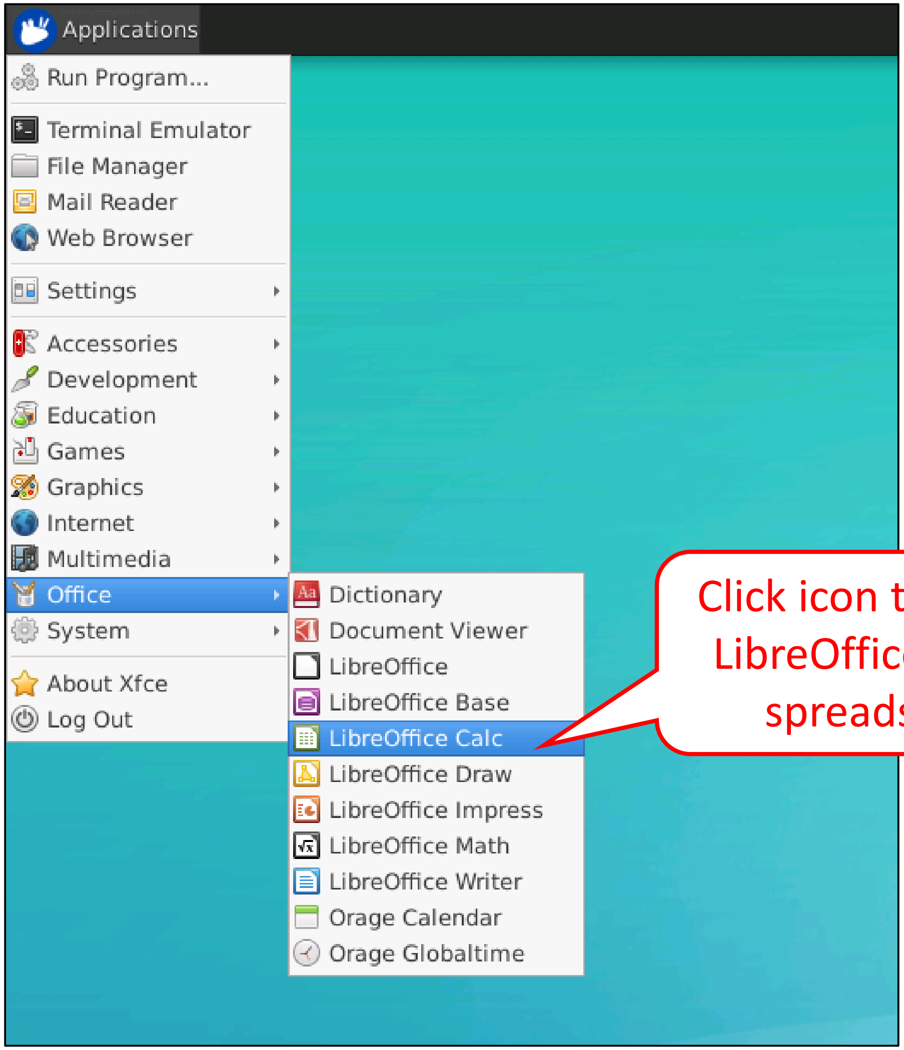
There will be no values in the blue cells

We keep adding difference columns until the **rightmost** column reaches a **steady** value

Creating a Spreadsheet

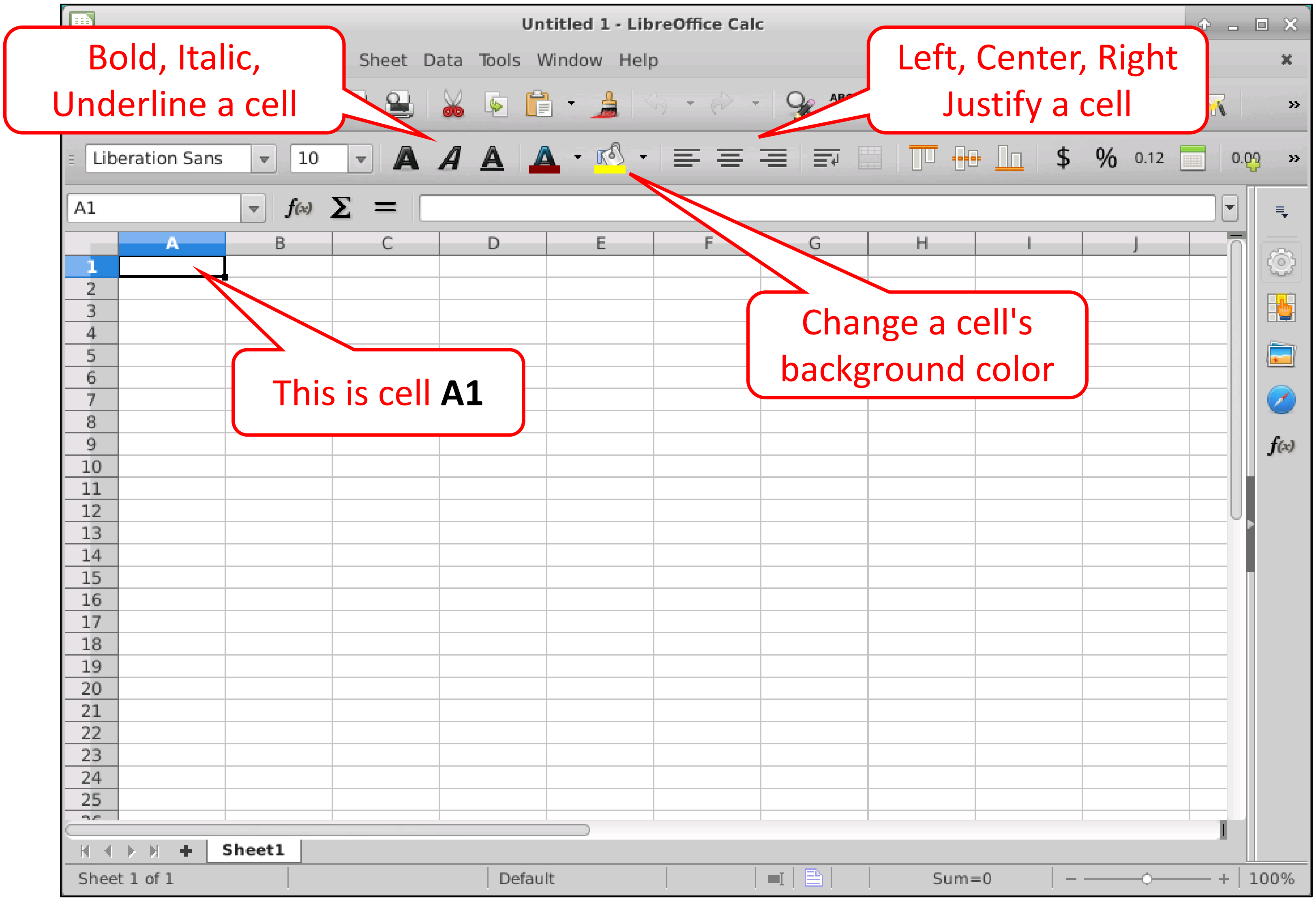
- A **spreadsheet** is a flexible computing tool that allows you to enter data and write formulas to operate on that data
- Everything is based on the concept of a "**cell**" that has a unique **column** (letter) and **row** (number) address
- A formula entered in one cell can reference data in one or more other cells by using cell addresses, or by using a **range of cell addresses**
- When source data cells are updated, the spreadsheet *automatically recalculates all dependent formula cells*
- Graphs can be created to depict the values in cells

Creating a Spreadsheet

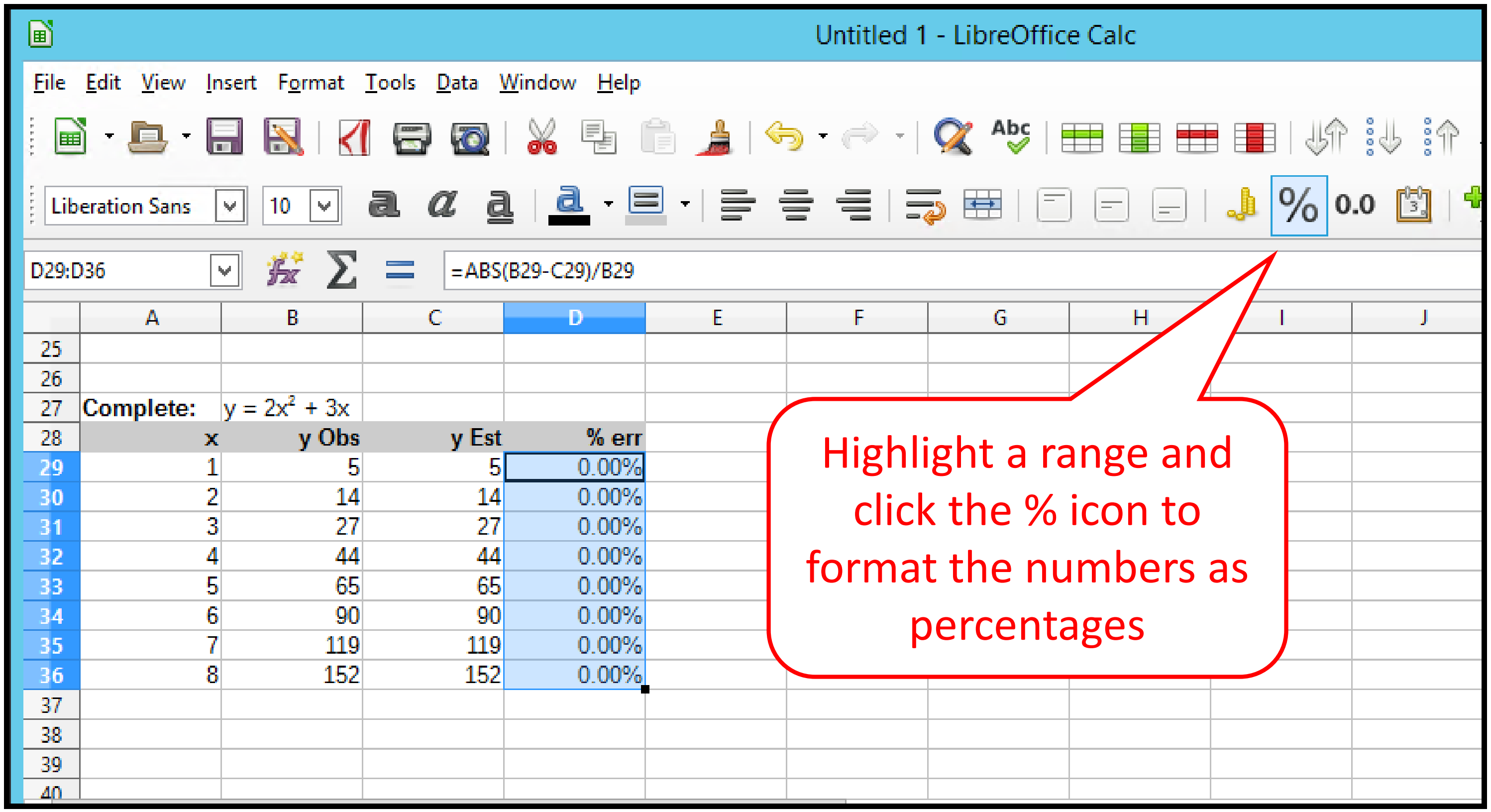


Click icon to launch LibreOffice “Calc” spreadsheet

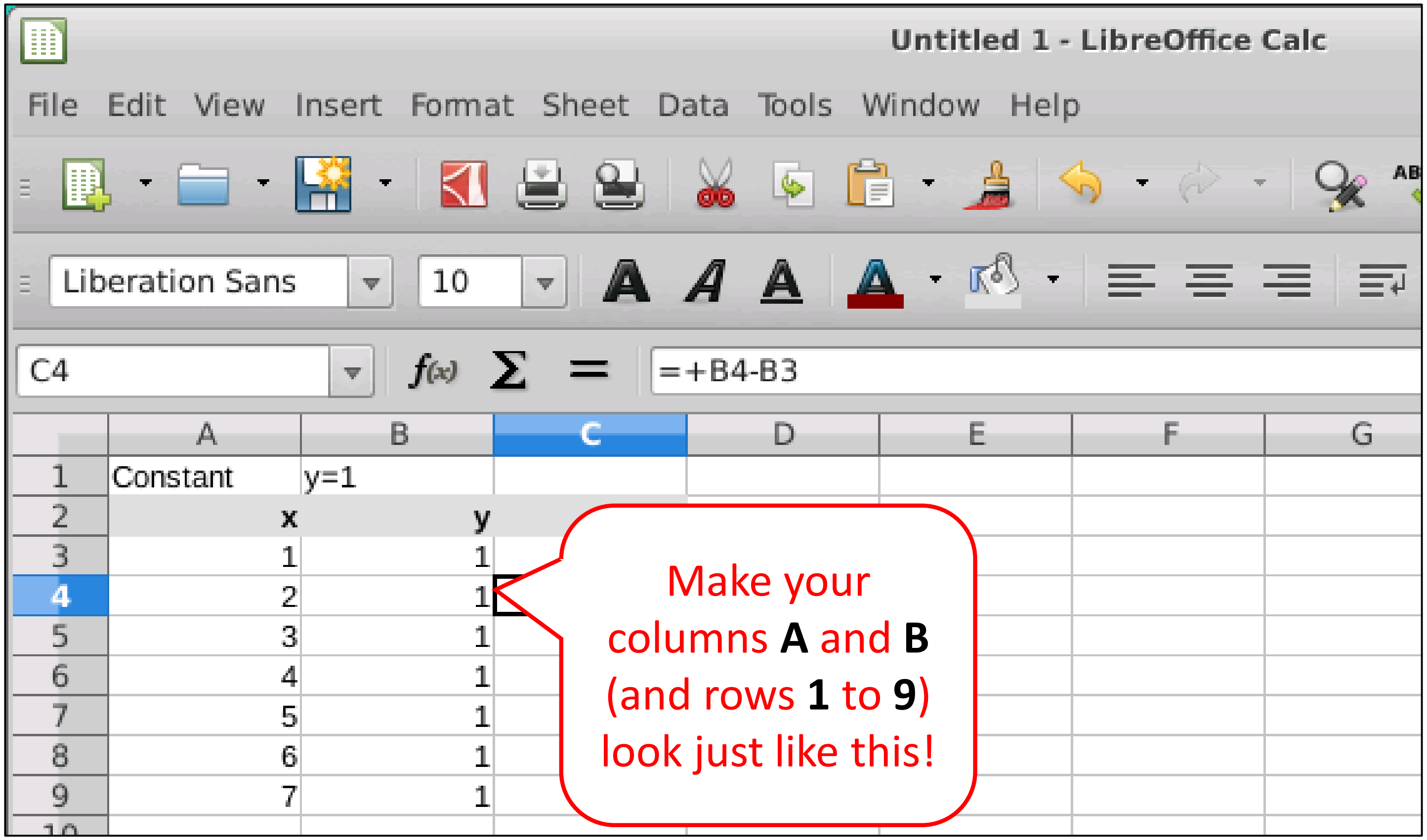
Lab 1 - Creating a Spreadsheet



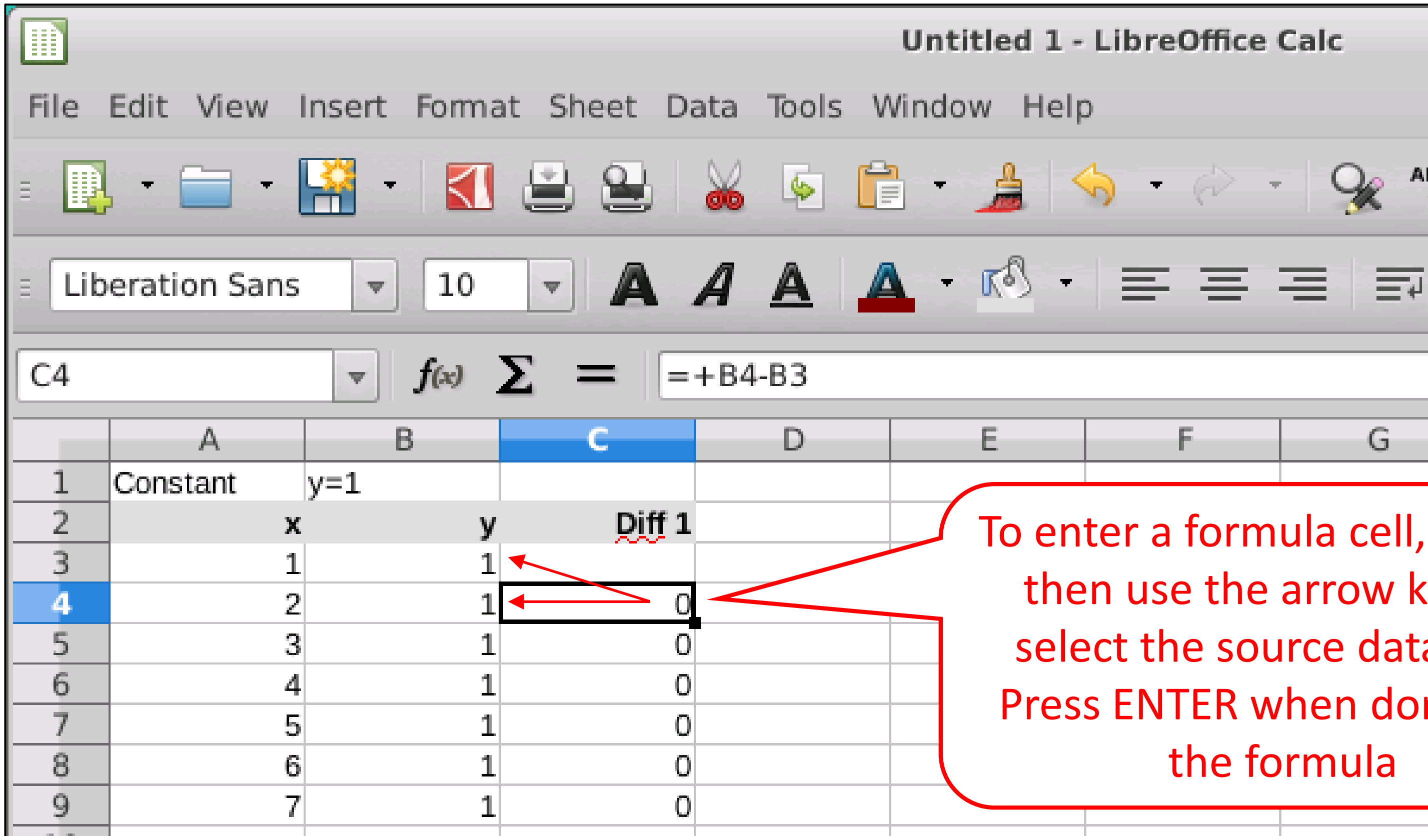
Changing a Cell's Number Format



Lab 1 – **Constant** Difference Table

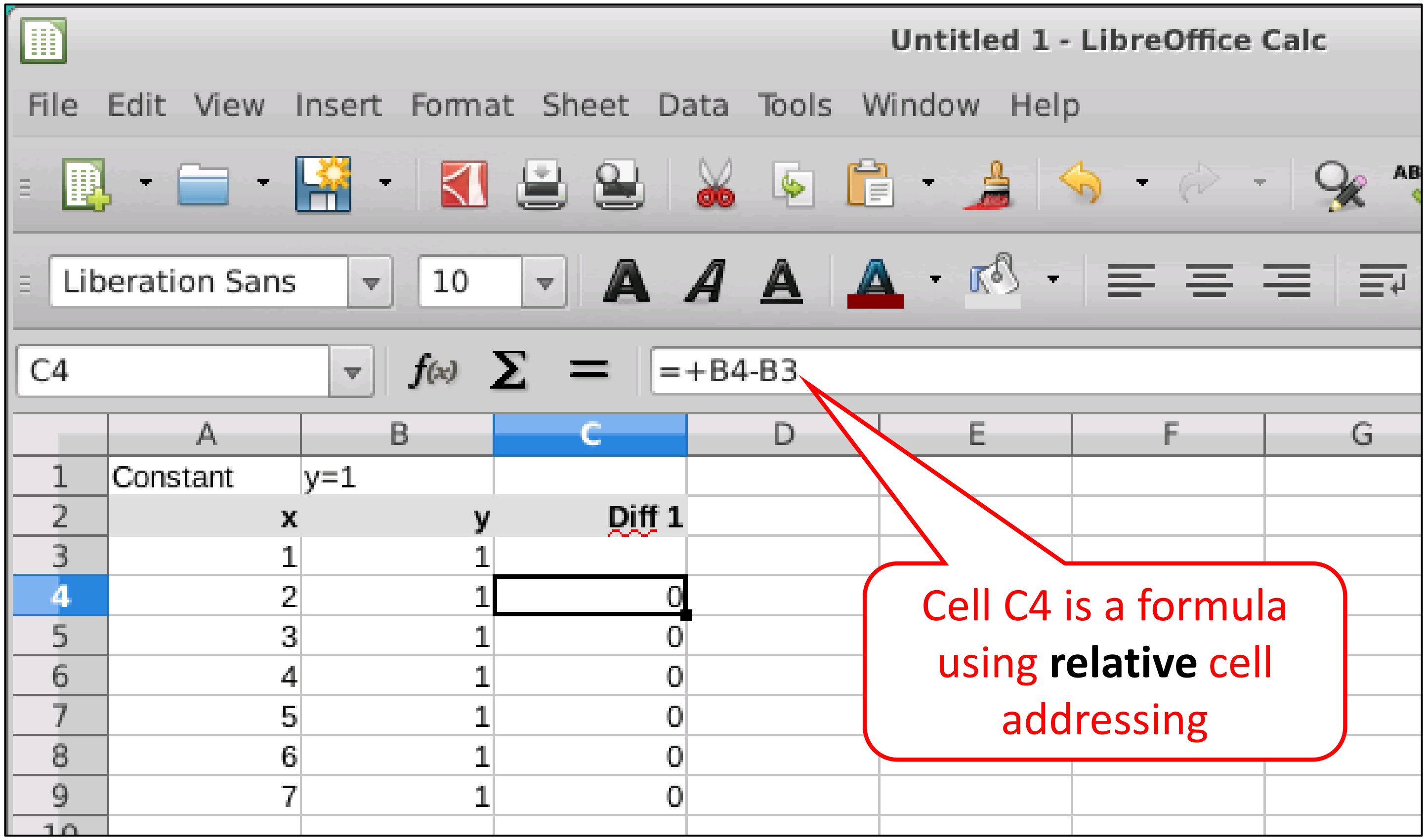


Lab 1 – **Constant** Difference Table

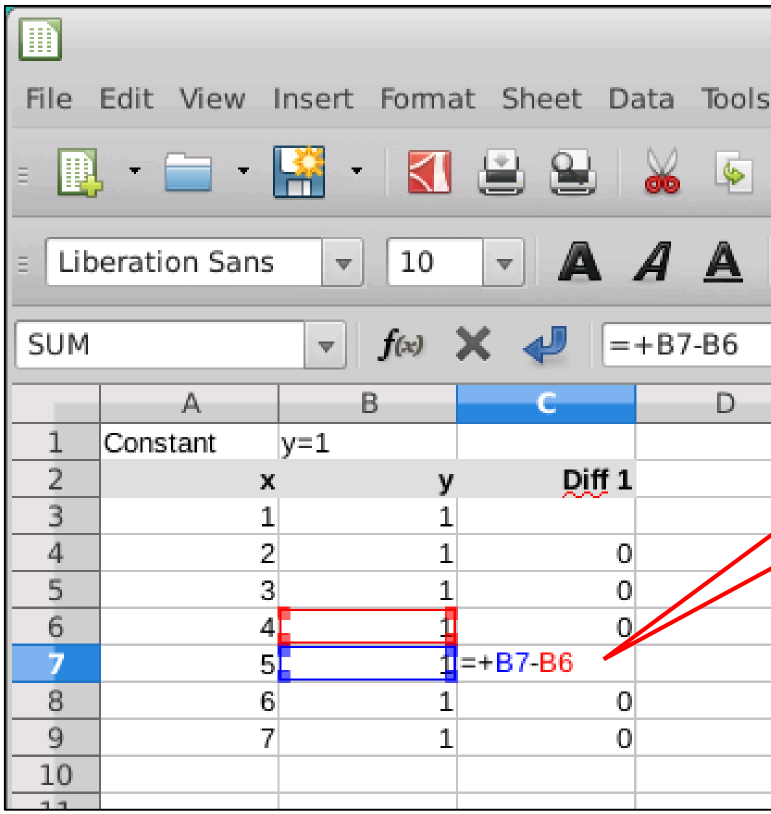


To enter a formula cell, press **+**, then use the arrow keys to select the source data cells. Press ENTER when done with the formula

Lab 1 – **Constant** Difference Table

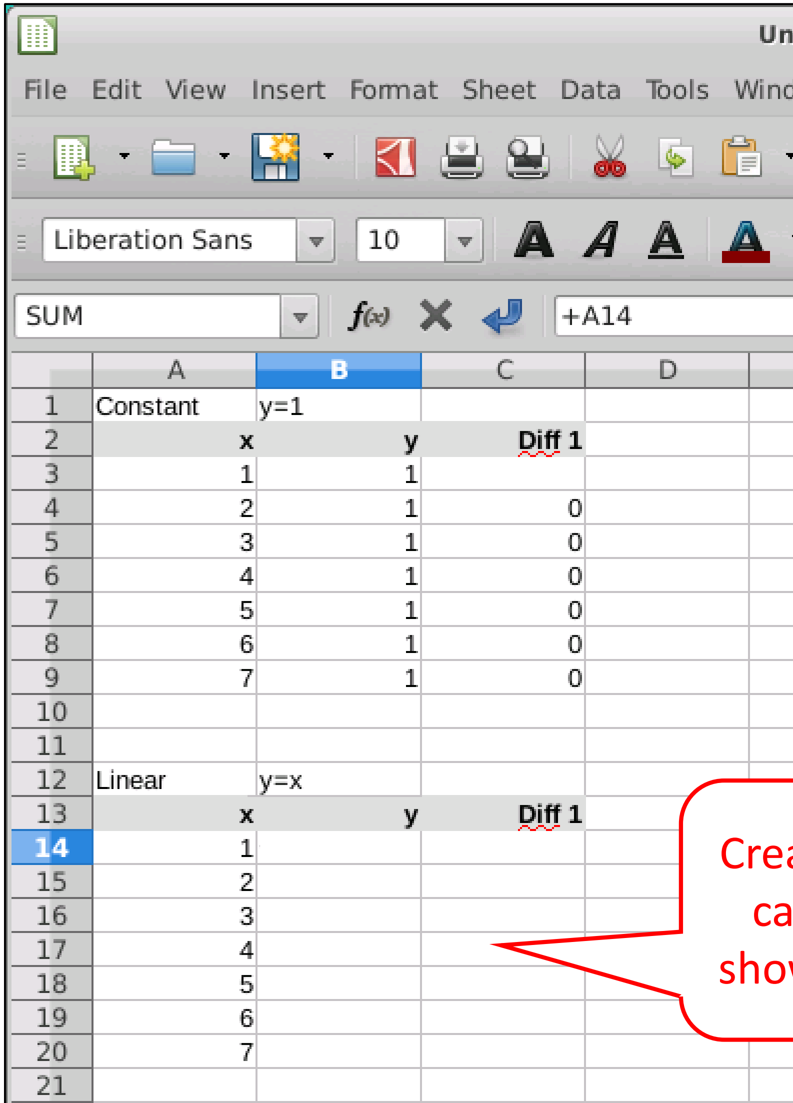


Lab 1 – **Constant** Difference Table



To check a formula cell, press **F2** to edit the formula – color coding then identifies the cells

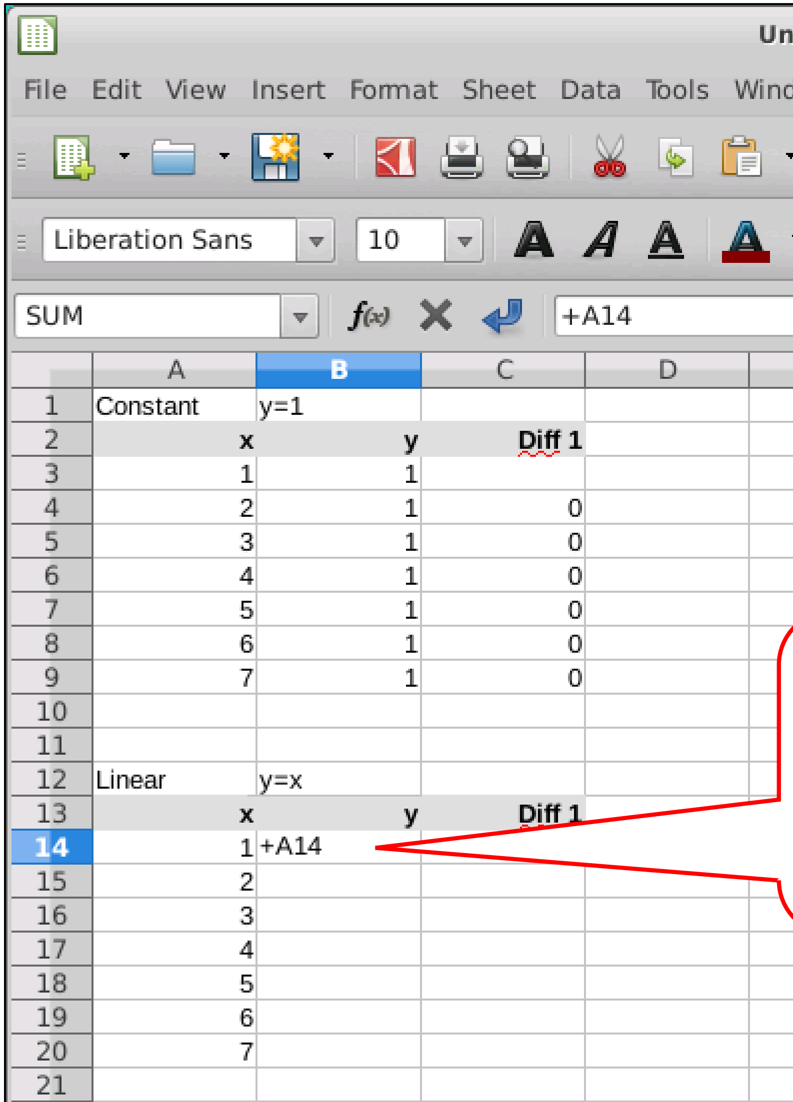
Lab 1 – **Linear** Difference Table



	A	B	C	D
1	Constant	y=1		
2	x	y	Diff 1	
3	1	1		
4	2	1	0	
5	3	1	0	
6	4	1	0	
7	5	1	0	
8	6	1	0	
9	7	1	0	
10				
11				
12	Linear	y=x		
13	x	y	Diff 1	
14	1			
15	2			
16	3			
17	4			
18	5			
19	6			
20	7			
21				

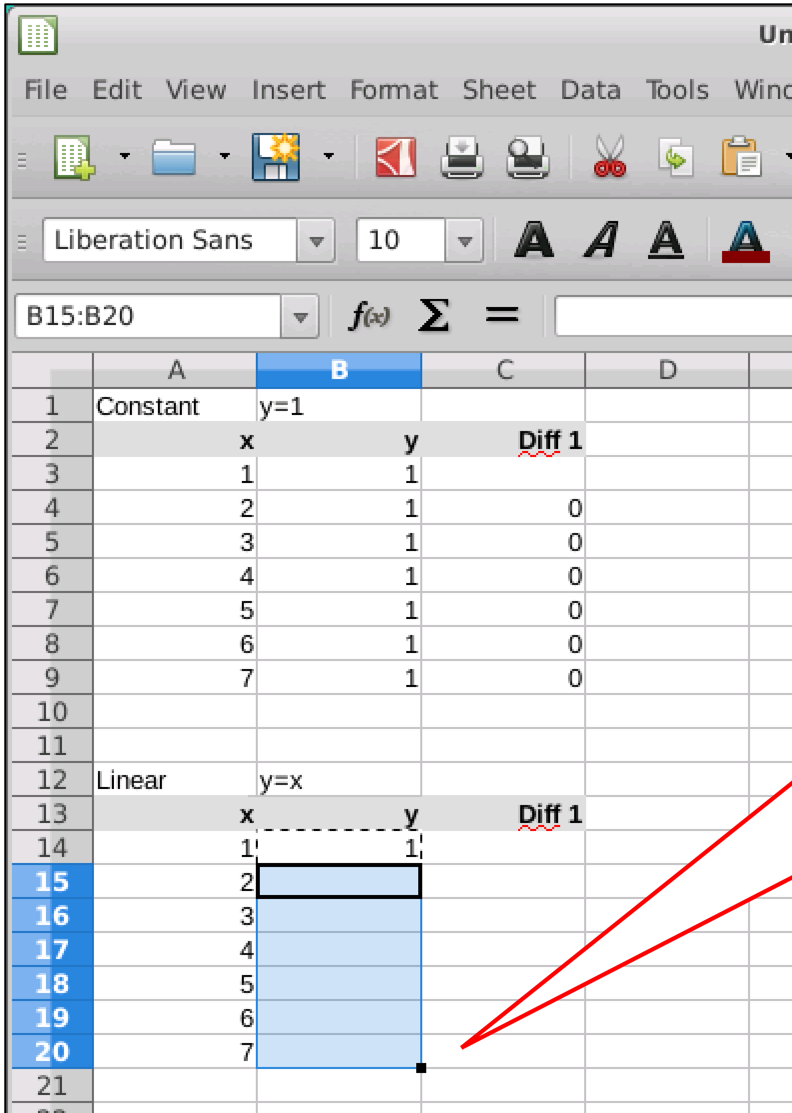
Create a new table called **Linear** as shown (3 columns)

Lab 1 – **Linear** Difference Table



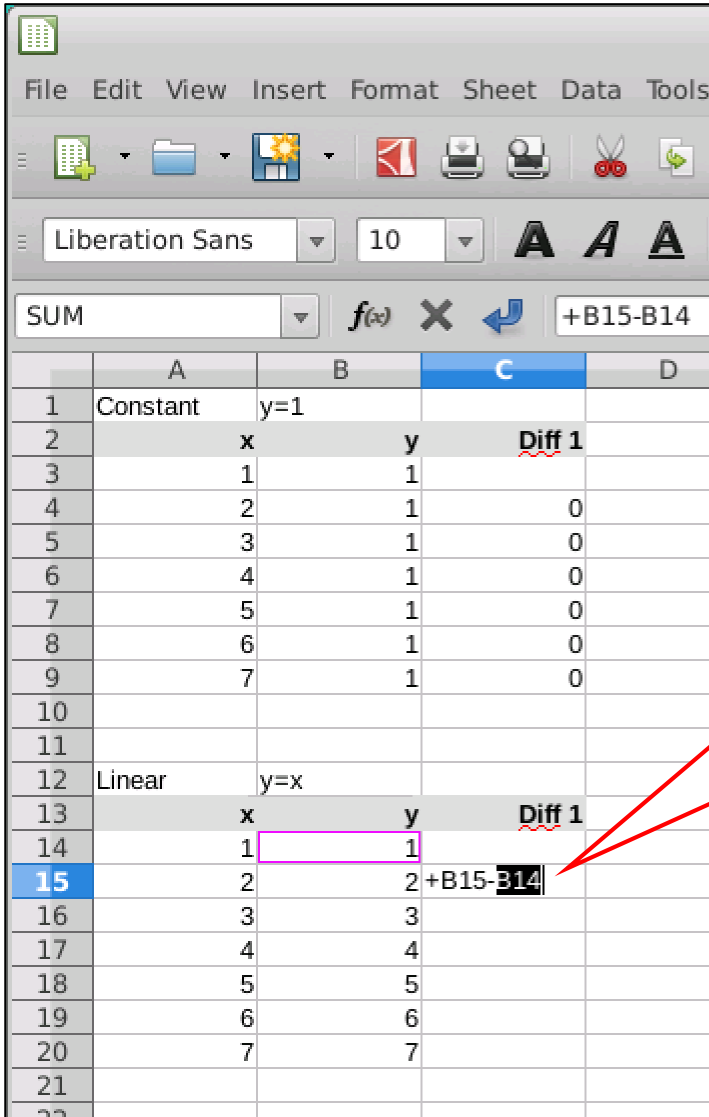
To enter a formula cell, press **+**, then use the arrow keys to select the source data cells. Press ENTER when done with the formula

Lab 1 – **Linear** Difference Table



To copy a formula to other cells, highlight the source cell, press "Control + C" to copy, then use SHIFT + arrow keys to highlight a destination **range**, and then press ENTER to paste the formula

Lab 1 – **Linear** Difference Table



	A	B	C	D
1	Constant	y=1		
2	x	y	Diff 1	
3	1	1		
4	2	1	0	
5	3	1	0	
6	4	1	0	
7	5	1	0	
8	6	1	0	
9	7	1	0	
10				
11				
12	Linear	y=x		
13	x	y	Diff 1	
14	1	1		
15	2	2	+B15-B14	
16	3	3		
17	4	4		
18	5	5		
19	6	6		
20	7	7		
21				
22				

Create the **Diff 1** column using the difference between the left-adjacent cell and the cell above that one, then copy that formula down the remaining cells in the **Diff 1** column.

Lab 1 – **Quadratic** Difference Table

22				
23	Quadratic	y=x^2		
24	**x**	**y**	**Diff 1**	**Diff 2**
25	1			
26	2			
27	3			
28	4			
29	5			
30	6			
31	7			
32				

Create a new table called **Quadratic** as shown (4 columns)

Lab 1 – **Quadratic** Difference Table

	A	B	C	D
22				
23	Quadratic	y=x^2		
24	x	y	Diff 1	Diff 2
25	1	+A25^2		
26	2			
27	3			
28	4			
29	5			
30	6			
31	7			
32				

Create the y column with the **exponent 2** and then copy down

	A	B	C	D
23	Quadratic	y=x^2		
24	x	y	Diff 1	Diff 2
25	1	1		
26	2	4	3	
27	3	9	5	+C27-C26
28	4	16	7	
29	5	25	9	
30	6	36	11	
31	7	49	13	

Add the **Diff 1** & **Diff 2** columns and formulas and then copy down

	A	B	C	D
23	Quadratic	y=x^2		
24	x	y	Diff 1	Diff 2
25	1	1		
26	2	4	3	
27	3	9	5	2
28	4	16	7	2
29	5	25	9	2
30	6	36	11	2
31	7	49	13	2
32				

The completed **quadratic** difference table

Lab 1 – **Cubic** and **Quartic** Tables

34	Cubic	y=x^3				
35	x	y	Diff 1	Diff 2	Diff 3	
36	1	1				
37	2	8	7			
38	3	27	19	12		
39	4	64	37	18	6	
40	5	125	61	24	6	
41	6	216	91	30	6	
42	7	343	127	36	6	
43						
44						
45	Quadratic	y=x^4				
46	x	y	Diff 1	Diff 2	Diff 3	Diff 4
47	1	1				
48	2	16	15			
49	3	81	65	50		
50	4	256	175	110	60	
51	5	625	369	194	84	24
52	6	1296	671	302	108	24
53	7	2401	1105	434	132	24
54						

Add the **Cubic** and **Quartic** Difference Tables

Difference Tables – The Big Picture

- A difference table calculates the delta between **successive** values (in the **range**) of a given function
- For every higher power of the *independent* variable (in the **domain**) we add **another** difference column
- Difference column #2 is the gap between successive values of difference column #1, etc.
- We keep adding difference columns until the value in the **rightmost** column is the same for every row – this is called achieving a **steady state**
- The **cubic** table needed **3** difference columns – the **quartic** table needed **4** difference columns to achieve **steady state**

Difference Tables – The Big Picture

34	Cubic	y=x^3				
35	x	y	Diff 1	Diff 2	Diff 3	
36	1	1				
37	2	8	7			
38	3	27	19	12		
39	4	64	37	18	6	
40	5	125	61	24	6	
41	6	216	91	30	6	
42	7	343	127	36	6	
43						
44						
45	Quadratic	y=x^4				
46	x	y	Diff 1	Diff 2	Diff 3	Diff 4
47	1	1				
48	2	16	15			
49	3	81	65	50		
50	4	256	175	110	60	
51	5	625	369	194	84	24
52	6	1296	671	302	108	24
53	7	2401	1105	434	132	24
54						

Keep adding difference columns until the **rightmost** column reaches a **steady** value

Difference Tables – What They Reveal

		Steady Difference
Constant	y = 1	0
Linear	y = x	1
Quadratic	$y = x^2$	2
Cubic	$y = x^3$	6
Quartic	$y = x^4$	24

Expected Steady Difference for $(1)x^{\boldsymbol{n}} = \boldsymbol{n}!$

$$(1)x^4 = 4! = 4 \times 3 \times 2 \times 1 = 24$$

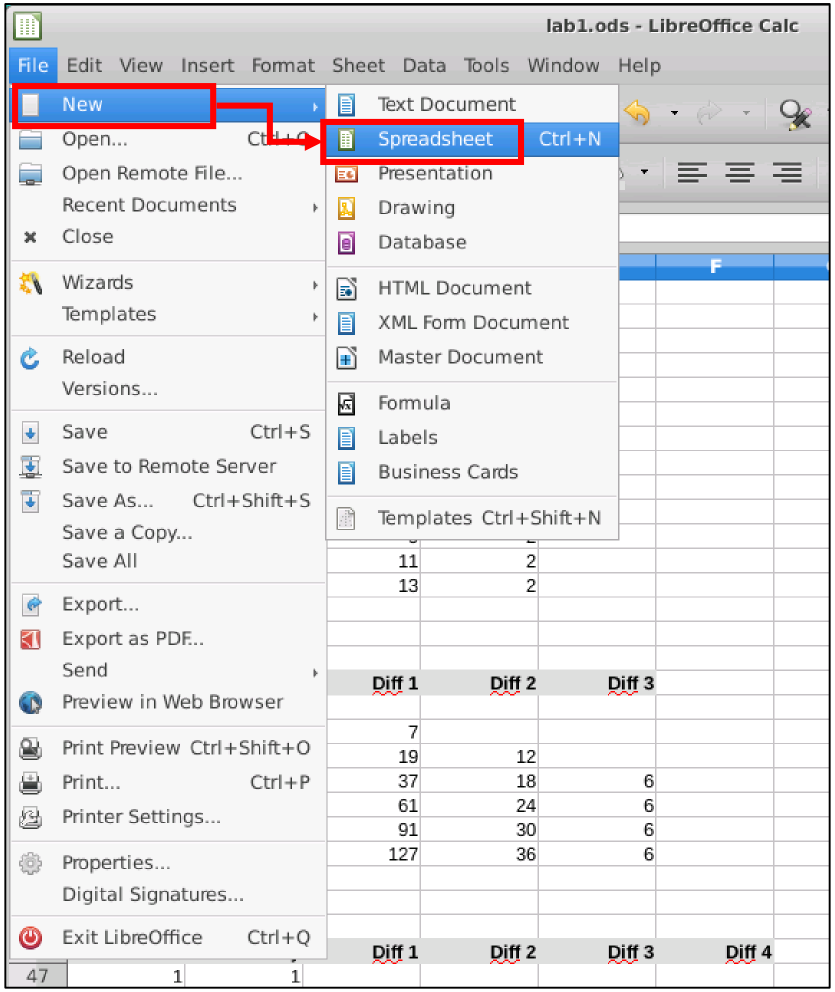
Difference Tables – What They Reveal

- If we believe a given a set of observations obeys an unknown general power law, we can use difference tables to systematically reveal this hidden functional **generator**
- We start by "guessing" the highest **power** of **x** that would likely be in the underlying generator
- We compare the *expected* steady state values for each power of x with the our observed values
- As we determine the **coefficient** for each decreasing power of **x**, we begin to expose (*one term at a time*) the underlying functional equation that generated the original sequence

Lab 2

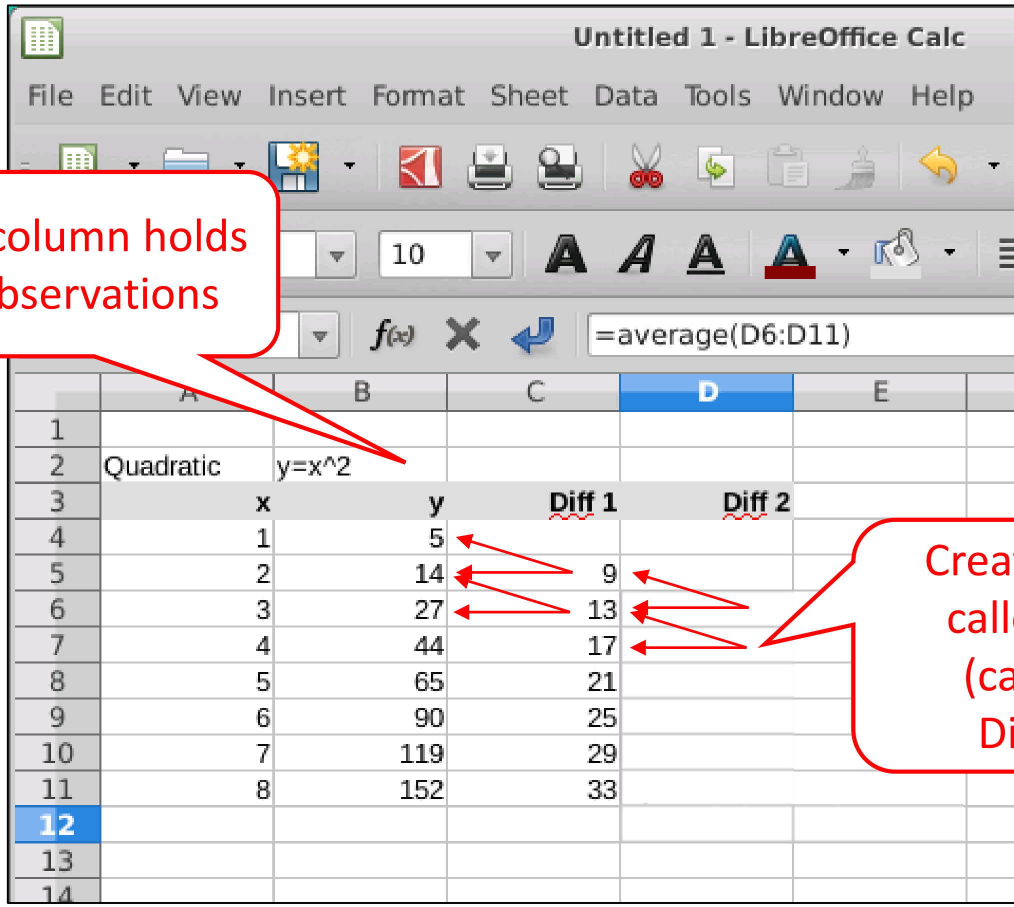
- Find an equation to generate this sequence: **5, 14, 27, 44, 65, 90, 119, 152 …**
- We will guess this data set is generated by a quadratic formula: $\boldsymbol{y = ax^2 + bx + c}$
- We have to figure out the values for $\boldsymbol{a, b, c}$
- We create a series of difference tables, working ***backwards***, starting with a **quadratic** table, then a **linear** table, and then a **constant** table (if necessary)
- We stop when our model produces **values that match our observations** to the desired level of ***accuracy***

Lab 2 – Create a new Spreadsheet



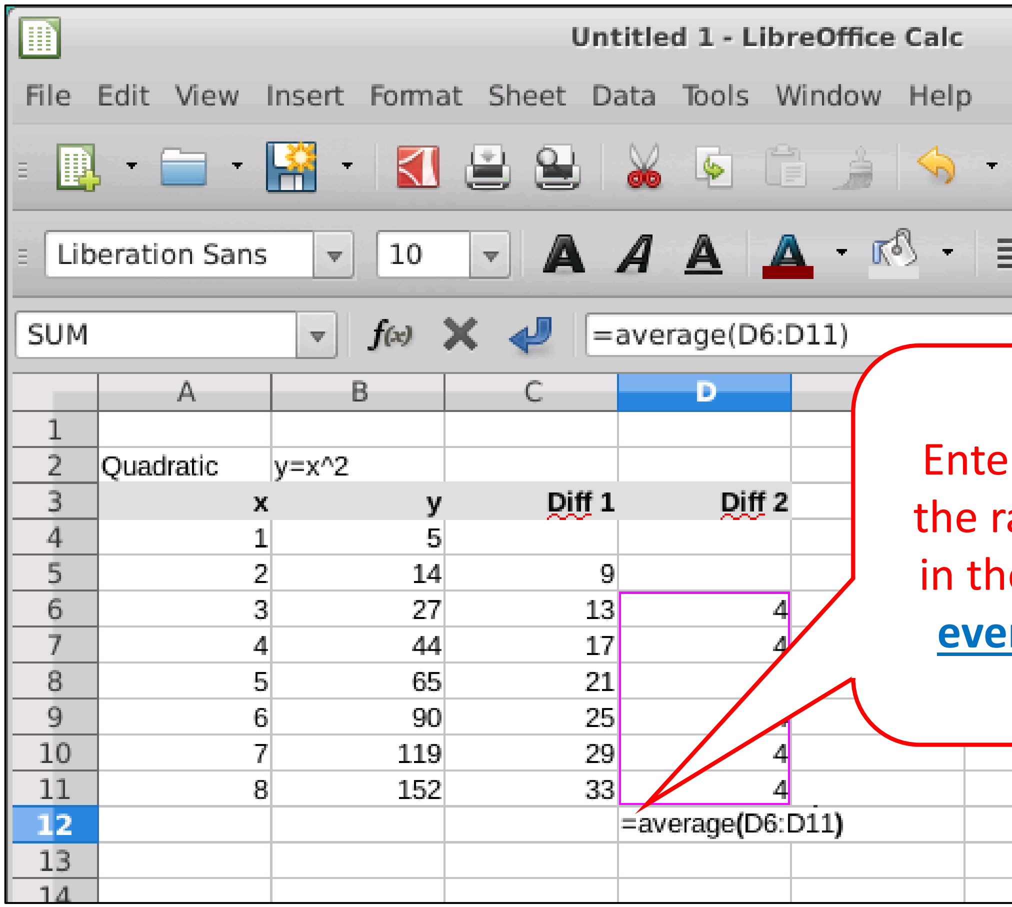
Lab 2 – **Quadratic** Difference Table

The **B** column holds the observations



Create a new table called **Quadratic** (calculate *both* Diff columns)

Calculating the Mean Steady Difference



	A	B	C	D
1				
2	Quadratic	y=x^2		
3	x	y	Diff 1	Diff 2
4	1	5		
5	2	14	9	
6	3	27	13	4
7	4	44	17	4
8	5	65	21	[illegible]
9	6	90	25	[illegible]
10	7	119	29	4
11	8	152	33	4
12				=average(D6:D11)
13				
14				

Enter **=average** then select the range of cells to include in the formula – **do this for <u>every</u> table from now on**

Calculating the Coefficient of each Term

2	Quadratic	y=x^2			
3	x	y	Diff 1	Diff 2	
4	1	5			
5	2	14	9		
6	3	27	13	4	
7	4	44	17	4	
8	5	65	21	4	
9	6	90	25	4	
10	7	119	29	4	
11	8	152	33	4	
12			mean:	4	
13					

- The $y = (1)x^2$ table had an *expected* steady difference of **2**
- But in our sequence we found a steady difference **mean** of **4**
- This means the coefficient for the x^2 term is $a = \frac{\mathbf{4}}{\mathbf{2}} = \mathbf{2}$
- So now we know the generator is $y \approx \mathbf{2}x^2$

Lab 2 – **Linear** Difference Table

15	Linear	y=x			
16	**x**	**y Obs**	**y Est**	**delta**	**Diff 1**
17	1	5			
18	2	14			
19	3	27			
20	4	44			
21	5	65			
22	6	90			
23	7	119			
24	8	152			
25					

Create a new table called **Linear**

Add the **y Est** and **delta** columns

The **y Obs** column holds the observed values

Delta = Observed - Expected

15	Linear	y=x			
16	x	y Obs	y Est	delta	Diff 1
17	1	5	=+2*A17^2		
18	2	14			
19	3	27			
20	4	44			
21	5	65			
22	6	90			
23	7	119			
24	8	152			

15	Linear	y=x			
16	x	y Obs	y Est	delta	Diff 1
17	1	5	2	=+B17-C17	
18	2	14			
19	3	27			
20	4	44			
21	5	65			
22	6	90			
23	7	119			
24	8	152			

- $y_{obs} = given\ values$
- $y_{est} = \mathbf{2}x^2$
- $\Delta = y_{obs} - y_{est}$

15	Linear	y=x			
16	x	y Obs	y Est	delta	Diff 1
17	1	5	2	3	
18	2	14	8	6	=+D18-D17
19	3	27	18	9	
20	4	44	32	12	
21	5	65	50	15	
22	6	90	72	18	
23	7	119	98	21	
24	8	152	128	24	

Add the **Diff 1** column and formula, then copy that down all rows

Delta = Observed - Expected

15	Linear	y=x				
16	x	y Obs	y Est	delta	Diff 1	
17	1	5	2	3		
18	2	14	8	6	3	
19	3	27	18	9	3	
20	4	44	32	12	3	
21	5	65	50	15	3	
22	6	90	72	18	3	
23	7	119	98	21	3	
24	8	152	128	24	3	
25				mean:	=+AVERAGE(E18:E24)	
26						

- $y_{obs} = given\ values$
- $y_{est} = 2x^2$
- $\Delta = y_{obs} - y_{est}$
- Observed mean steady value = **3**
- *Expected* **linear** steady value = **1**
- $\therefore b = \frac{\mathbf{3}}{\mathbf{1}} = \mathbf{3}$
- Generator is now $y \approx 2x^2 + \mathbf{3}x$

Create a new table called **Complete**

% Relative Error

28	Complete	y=2*x^2+3*x			
29	x	y Obs	y Est	% Err	
30	1	5	=+2*A30^2+3*A30		
31	2	14	14	0.00%	
32	3	27	27	0.00%	
33	4	44	44	0.00%	
34	5	65	65	0.00%	
35	6	90	90	0.00%	
36	7	119	119	0.00%	
37	8	152	152	0.00%	
38					

- ABS = Absolute Value
- $\%_{err} = \left|\frac{(y_{obs} - y_{est})}{y_{obs}}\right|$
- 0% error = perfect fit
- $y = 2x^2 + 3x$

$2x^2 + 3x$ generates

5, 14, 27, 44, 65, 90, 119, 152 …

28	Complete	y=2*x^2+3*x			
29	x	y Obs	y Est	% Err	
30	1	5	5	=+(B30-C30)/B30	
31	2	14	14	0.00%	
32	3	27	27	0.00%	
33	4	44	44	0.00%	
34	5	65	65	0.00%	
35	6	90	90	0.00%	
36	7	119	119	0.00%	
37	8	152	152	0.00%	
38					

Lab 3

- Create a new Lab 3 spreadsheet and generate the difference tables to find the underlying equation that generate this sequence: **36, 103, 244, 489, 868, 1411, 2148, 3109, 4324…**
- Hint: This data set is generated by a cubic formula:

$$\boldsymbol{y = ax^3 + bx^2 + cx + d}$$

- You have to figure out the values for $\boldsymbol{a, b, c, d}$
- Create difference tables, starting with a **cubic**, then **quadratic**, then **linear**, then **constant** table (if necessary)
- Stop when your model produces **values that match our observations** with 0% relative error
- Create a new worksheet named **Lab 3.ods**

Lab 3 – **Cubic** Difference Table

	A	B	C	D	E
1	Cubic	y=x^3			
2	x	y Obs	Diff 1	Diff 2	Diff 3
3	1	36			
4	2	103	67		
5	3	244	141	74	
6	4	489	245	104	30
7	5	868	379	134	30
8	6	1411	543	164	30
9	7	2148	737	194	30
10	8	3109	961	224	30
11	9	4324	1215	254	30
12				mean:	30

Create a new table called **Cubic** (calculate *all three* Diff columns)

		Steady Difference
Constant	y = 1	0
Linear	y = x	1
Quadratic	$y = x^2$	2
Cubic	$y = x^3$	6
Quartic	$y = x^4$	[illegible]

Expected difference for a cubic = **6**, while observed steady diff <u>mean</u> = **30**, so coefficient of x^3 must be

$$\frac{\mathbf{30}}{\mathbf{6}} = \mathbf{5}$$

So far … $y = \mathbf{5}x^3$

Lab 3 – **Quadratic** Difference Table

	A	B	C	D	E	F
1	Cubic	y=x^3				
2	x	y Obs	Diff 1	Diff 2	Diff 3	
3	1	36				
4	2	103	67			
5	3	244	141	74		
6	4	489	245	104	30	
7	5	868	379	134	30	
8	6	1411	543	164	30	
9	7	2148	737	194	30	
10	8	3109	961	224	30	
11	9	4324	1215	254	30	
12				mean:	30	
13						
14	Quadratic	y=x^2				
15	x	y Obs	y Est	delta	Diff 1	Diff 2
16	1	36	=+5*A16^3	31		
17	2	103	40	63	32	
18	3	244	135	109	46	14
19	4	489	320	169	60	14
20	5	868	625	243	74	14
21	6	1411	1080	331	88	14
22	7	2148	1715	433	102	14
23	8	3109	2560	549	116	14
24	9	4324	3645	679	130	14
25					mean:	14
26						

Create a new table called **Quadratic**

Add the **y Est, delta,** and *both* **Diff** columns

So far … $y = \mathbf{5}x^3$

Lab 3 – **Quadratic** Difference Table

	A	B	C	D	E	F
1	Cubic	y=x^3				
2	x	y Obs	Diff 1	Diff 2	Diff 3	
3	1	36				
4	2	103	67			
5	3	244	141	74		
6	4	489	245	104	30	
7	5	868	379	134	30	
8	6	1411	543	164	30	
9	7	2148	737	194	30	
10	8	3109	961	224	30	
11	9	4324	1215	254	30	
12				mean:	30	
13						
14	Quadratic	y=x^2				
15	x	y Obs	y Est	delta	Diff 1	Diff 2
16	1	36	=+5*A16^3	31		
17	2	103	40	63	32	
18	3	244	135	109	46	14
19	4	489	320	169	60	14
20	5	868	625	243	74	14
21	6	1411	1080	331	88	14
22	7	2148	1715	433	102	14
23	8	3109	2560	549	116	14
24	9	4324	3645	679	130	14
25					mean:	14
26						

		Steady Difference
Constant	$y = 1$	0
Linear	$y = x$	1
Quadratic	$y = x^2$	2
Cubic	$y = x^3$	
Quartic	$y = x^4$	

Expected difference for a quadratic = **2**, while observed steady diff <u>mean</u> = **14**, so coefficient of x^2 must be

$$\frac{\mathbf{14}}{\mathbf{2}} = \mathbf{7}$$

Now so far … $y \approx \mathbf{5}x^3 + \mathbf{7}x^2$

Lab 3 – **Linear** Difference Table

Remember to add the new quadratic term to the **estimated y** function

27	Linear	y=x			
28	x	y Obs	y Est	delta	Diff 1
29	1	36	=+5*A29^3+7*A29^2		
30	2	103	68	35	11
31	3	244	198	46	11
32	4	489	432	57	11
33	5	868	800	68	11
34	6	1411	1332	79	11
35	7	2148	2058	90	11
36	8	3109	3008	101	11
37	9	4324	4212	112	11
38				mean:	11
39					

		Steady Difference
Constant	$y = 1$	0
Linear	$y = x$	1
Quadratic	$y = x^2$	2
Cubic	$y = x^3$	
Quartic	$y = x^4$	

Steady difference for **linear** = **1**, while observed steady diff mean = **11**, so coefficient of x must be

$$\frac{11}{1} = 11$$

Now so far … $y \approx \mathbf{5}x^3 + \mathbf{7}x^2 + \mathbf{11}x$

Lab 3 – **Constant** Difference Table

40	Constant	y=c			
41		**x**	**y Obs**	**y Est**	**delta**
42		1	36	23	13
43		2	103	90	13
44		3	244	231	13
45		4	489	476	13
46		5	868	855	13
47		6	1411	1398	13
48		7	2148	2135	13
49		8	3109	3096	13
50		9	4324	4311	13
51				**mean:**	13
52					
53	Complete				
54		**x**	**y Obs**	**y Est**	**% Error**
55		1	36	=+5*A55^3+7*A55^2+11*A55+13	
56		2	103	103	0.00%
57		3	244	244	0.00%
58		4	489	489	0.00%
59		5	868	868	0.00%
60		6	1411	1411	0.00%
61		7	2148	2148	0.00%
62		8	3109	3109	0.00%
63		9	4324	4324	0.00%
64					

If the delta values are all the same, then that value is the final **constant** term

Functional Equation

$$y = 5x^3 + 7x^2 + 11x + 13$$

$$with\ x \in \mathbb{Z}^+$$

Generates the sequence:
36, 103, 244, 489, 868, 1411, 2148, 3109, 4324…

Lab 4

Elapsed Time (10 mins each)	Observed Counter
1	205
2	430
3	677
4	945
5	1233
6	1542
7	1872
8	2224

- This case study will use data from an experiment that measured the **counter** on a tape machine vs. the elapsed **time** the tape was played
- Ideally there would be a *linear* relationship between these two variables
- But due to the constantly changing circumference of a tape reel as it is played, the relationship is **non-linear**

Lab 4

Elapsed Time (10 mins each)	Observed Counter
1	205
2	430
3	677
4	945
5	1233
6	1542
7	1872
8	2224

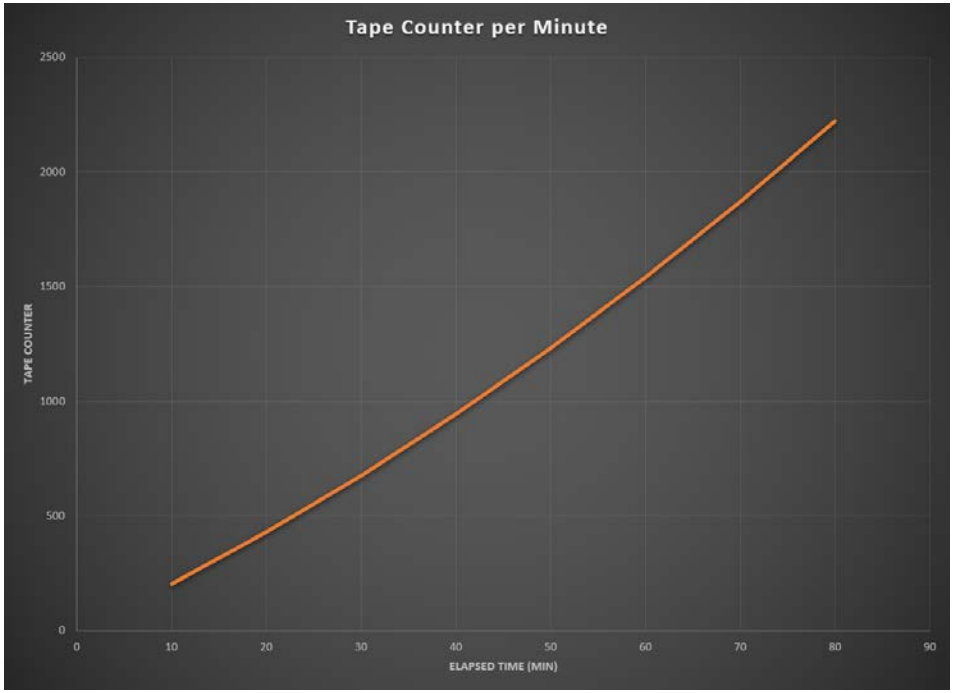
- Find an equation to model the tape counter as a function of playing time
  - **X** is the number of 10 minute blocks the tape has been playing from the beginning (1…8)
  - **Y** is the counter on the tape player (linear feet)
- With a precise model we can answer this question:

  Where should we stop the tape to be exactly **65** minutes into the recording? $(x = 6.5)$

Lab 4

Elapsed Time (10 mins each)	Observed Counter
1	205
2	430
3	677
4	945
5	1233
6	1542
7	1872
8	2224

$$Assume$$
$$y = ax^2 + bx + c$$



Lab 4 – Difference Tables

1	Quadratic	y=x^2			
2		**x**	**y Obs**	**Diff 1**	**Diff 2**
3		1	205		
4		2	430	225	
5		3	677	247	22
6		4	945	268	21
7		5	1233	288	20
8		6	1542	309	21
9		7	1872	330	21
10		8	2224	352	22
11				**mean:**	**21.1667**

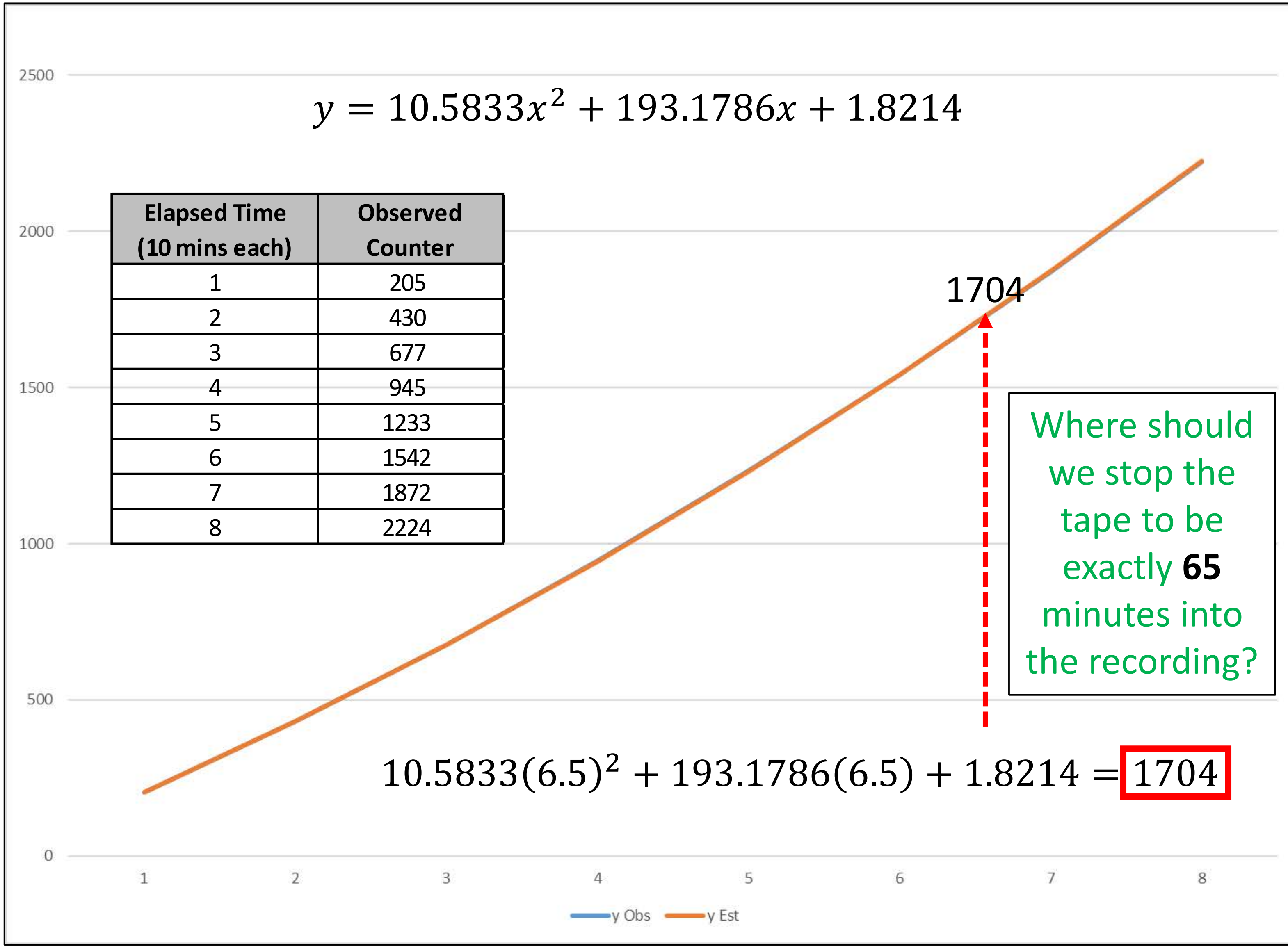
13	Linear	y=x					
14		**x**	**y Obs**	**y Est**	**delta**	**Diff 1**
15		1	205	10.5833	194.4167	
16		2	430	42.3333	387.6667	193.2500
17		3	677	95.2500	581.7500	194.0833
18		4	945	169.3333	775.6667	193.9167
19		5	1233	264.5833	968.4167	192.7500
20		6	1542	381.0000	1,161.0000	192.5833
21		7	1872	518.5833	1,353.4167	192.4167
22		8	2224	677.3333	1,546.6667	193.2500
23					**mean:**	**193.1786**

25	Constant	y=c			
26		**x**	**y Obs**	**y Est**	**delta**
27		1	205	203.7619	1.2381
28		2	430	428.6905	1.3095
29		3	677	674.7857	2.2143
30		4	945	942.0476	2.9524
31		5	1233	1,230.4762	2.5238
32		6	1542	1,540.0714	1.9286
33		7	1872	1,870.8333	1.1667
34		8	2224	2,222.7619	1.2381
35				**mean:**	**1.8214**

38	Complete				
39		**x**	**y Obs**	**y Est**	**% Err**
40		1	205	205.5833	-0.2846%
41		2	430	430.5119	-0.1190%
42		3	677	676.6071	0.0580%
43		4	945	943.8690	0.1197%
44		5	1233	1,232.2976	0.0570%
45		6	1542	1,541.8929	0.0069%
46		7	1872	1,872.6548	-0.0350%
47		8	2224	2,224.5833	-0.0262%

$$y_{est} = 10.5833x^2 + 193.1786x + 1.8214$$

Lab 4 – Model Predictions



$$y = 10.5833x^2 + 193.1786x + 1.8214$$

Elapsed Time (10 mins each)	Observed Counter
1	205
2	430
3	677
4	945
5	1233
6	1542
7	1872
8	2224

Where should we stop the tape to be exactly **65** minutes into the recording?

$$10.5833(6.5)^2 + 193.1786(6.5) + 1.8214 = \boxed{1704}$$

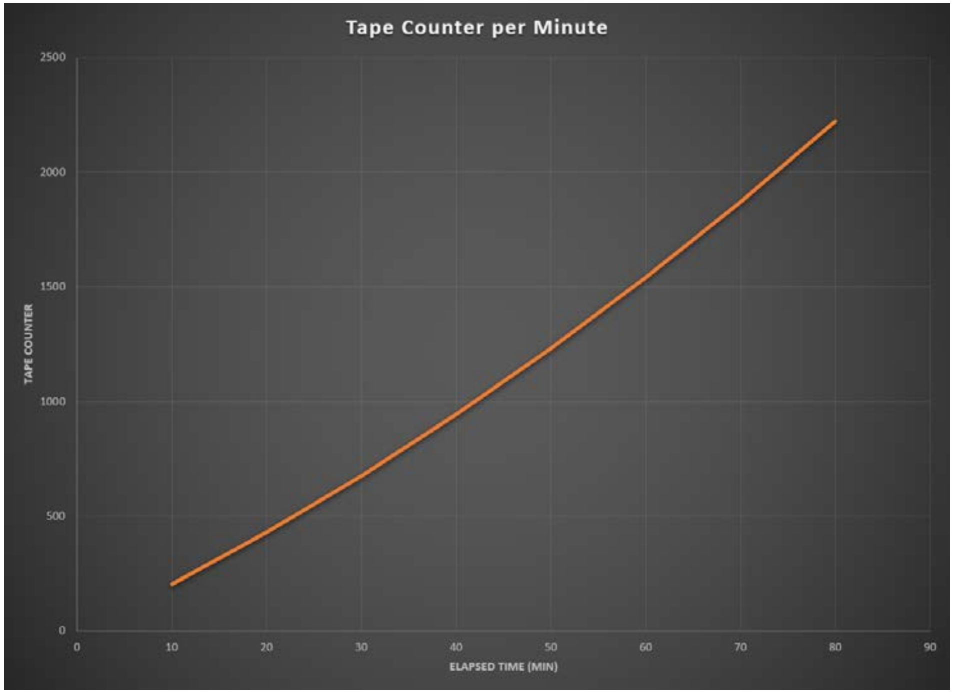
Lab 5 - Method of Least Squares

- Developed by Gauss, least squares finds the coefficients of a polynomial of a given degree that **approximates** a set of observations with **minimal variance** from the data
- The modeler selects a curve whose **general shape** matches the trend of the data
- The **partial derivatives** of the chosen polynomial are then determined, leading to a system of *linear* equations that can be solved using matrices and **Cramer's Rule**
- The coefficients of each term of the model polynomial are then determined, allowing us to estimate the unknown function value at any point within the given **domain**

Lab 5 - Method of Least Squares

$$Assume$$
$$y = ax^2 + bx + c$$

Elapsed Time (10 mins each)	Observed Counter
1	205
2	430
3	677
4	945
5	1233
6	1542
7	1872
8	2224



Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n} \left[y_i - \left(a x_i^2 + b x_i + c \right) \right]^2$$

$$(y - ax^2 - bx - c)(y - ax^2 - bx - c)$$

$$\begin{array}{l}
y^2 - ax^2y - bxy - cy \\
\quad -ax^2y \qquad\qquad + a^2x^4 + abx^3 + acx^2 \\
\qquad\quad -bxy \qquad\qquad + abx^3 \qquad\qquad + b^2x^2 + bcx \\
\qquad\qquad\quad -cy \qquad\qquad\quad + acx^2 \qquad\qquad + bcx + c^2
\end{array}$$

$$\boxed{S = y^2 - 2ax^2y - 2bxy - 2cy + a^2x^4 + 2abx^3 + 2acx^2 + b^2x^2 + 2bcx + c^2}$$

Note: we will set aside the summation operator for now…

Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n} \left[y_i - \left(ax_i^2 + bx_i + c\right)\right]^2$$

$$S = y^2 - 2ax^2y - 2bxy - 2cy + a^2x^4 + 2abx^3 + 2acx^2 + b^2x^2 + 2bcx + c^2$$

For S to have a **minimum**, these partial derivatives must exist:

$$\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0, \frac{\partial S}{\partial c} = 0$$

$$\frac{\partial S}{\partial a} = -2x^2y + 2ax^4 + 2bx^3 + 2cx^2 = 0$$

$$\frac{\partial S}{\partial b} = -2xy + 2ax^3 + 2bx^2 + 2cx = 0$$

$$\frac{\partial S}{\partial c} = -2y + 2ax^2 + 2bx + 2c = 0$$

Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n}\left[y_i - \left(ax_i^2 + bx_i + c\right)\right]^2$$

Divide through by the **GCD** of all terms

Move all terms <u>not</u> containing a, b, or c to the **RHS**

$$\frac{\partial S}{\partial a} = -2x^2y + 2ax^4 + 2bx^3 + 2cx^2 = 0$$

$$\boxed{(x^4)a + (x^3)b + (x^2)c = x^2y}$$

$$\frac{\partial S}{\partial b} = -2xy + 2ax^3 + 2bx^2 + 2cx = 0$$

$$\boxed{(x^3)a + (x^2)b + (x)c = xy}$$

$$\frac{\partial S}{\partial c} = -2y + 2ax^2 + 2bx + 2c = 0$$

$$\boxed{(x^2)a + (x)b + c = y}$$

Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n}\left[y_i - \left(ax_i^2 + bx_i + c\right)\right]^2$$

$$\sum_{i=1}^{n}(1)c = (n)c$$

Reintroduce the $\sum$ operators

But these $\sum$ values are just sums of the **known** (observed) values!

$$\boxed{\begin{aligned}\left(\sum x_i^4\right)a + \left(\sum x_i^3\right)b + \left(\sum x_i^2\right)c \\ \left(\sum x_i^3\right)a + \left(\sum x_i^2\right)b + \left(\sum x_i\right)c \\ \left(\sum x_i^2\right)a + \left(\sum x_i\right)b + (n)c\end{aligned}} \begin{aligned} = \\ = \\ = \end{aligned} \boxed{\begin{aligned}\left(\sum x_i^2 y_i\right) \\ \left(\sum x_i y_i\right) \\ \left(\sum y_i\right)\end{aligned}}$$

Pretend all the $\sum$ values where just some numbers....

Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n} \left[y_i - \left(a x_i^2 + b x_i + c\right)\right]^2$$

This is a system of 3 linear equations and 3 unknowns!

$$(4)a + \quad (5)b + (-2)c = (-14)$$

$$(7)a + (-1)b + \quad (2)c = \ (42)$$

$$(3)a + \quad (1)b + \ (4)c = \ (28)$$

Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n} \left[y_i - \left(ax_i^2 + bx_i + c\right)\right]^2$$

This is a system of 3 linear equations and 3 unknowns!

$$\boxed{\begin{aligned} \left(\sum x_i^4\right)a + \left(\sum x_i^3\right)b + \left(\sum x_i^2\right)c \\ \left(\sum x_i^3\right)a + \left(\sum x_i^2\right)b + \left(\sum x_i\right)c \\ \left(\sum x_i^2\right)a + \left(\sum x_i\right)b + \quad (n)c \end{aligned}} \begin{aligned} = \\ = \\ = \end{aligned} \boxed{\begin{aligned} \left(\sum x_i^2 y_i\right) \\ \left(\sum x_i y_i\right) \\ \left(\sum y_i\right) \end{aligned}}$$

*See Session **10**:* **Coefficient Matrix** **Value Vector**

Open Lab 5 - Method of Least Squares

$$Minimize\ S = \sum_{i=1}^{n}\left[y_i - (ax_i^2 + bx_i + c)\right]^2$$

$$\left(\sum x_i^4\right)a + \left(\sum x_i^3\right)b + \left(\sum x_i^2\right)c = \left(\sum x_i^2 y_i\right)$$

$$\left(\sum x_i^3\right)a + \left(\sum x_i^2\right)b + \left(\sum x_i\right)c = \left(\sum x_i y_i\right)$$

$$\left(\sum x_i^2\right)a + \left(\sum x_i\right)b + (n)c = \left(\sum y_i\right)$$

```
double sumX  = PowerSum(vecX, 1);
double sumX2 = PowerSum(vecX, 2);
double sumX3 = PowerSum(vecX, 3);
double sumX4 = PowerSum(vecX, 4);

double sumY  = PowerSum(vecY, 1);

double sumXY  = PowerSum(vecX, 1, vecY);
double sumX2Y = PowerSum(vecX, 2, vecY);

double coeffMatrix[3][3]{
    {sumX4, sumX3, sumX2},
    {sumX3, sumX2, sumX},
    {sumX2, sumX, 8}};

double valueVector[3]{sumX2Y, sumXY, sumY};
```

We have **8** sample points in the Lab 5 data set

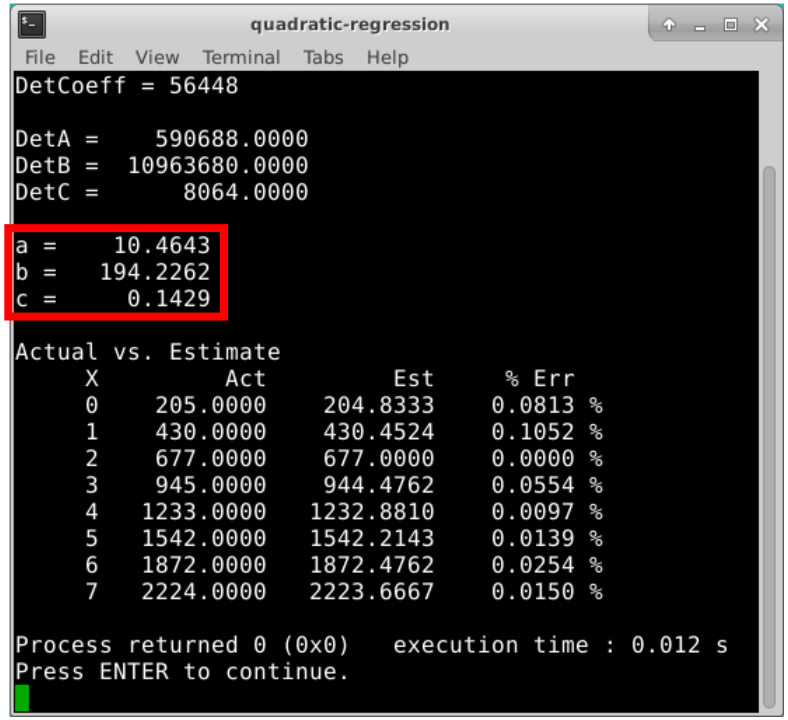
View Lab 5 - Method of Least Squares

```
double a = detA / detCoeff;
double b = detB / detCoeff;
double c = detC / detCoeff;

cout << "a = " << setw(10) << a << endl;
cout << "b = " << setw(10) << b << endl;
cout << "c = " << setw(10) << c << endl;
cout << endl;

cout << "Actual vs. Estimate" << endl;
cout << setw(6) << "X";
cout << setw(12) << "Act";
cout << setw(12) << "Est";
cout << setw(10) << "% Err";
cout << endl;

for (int i{}; i < 8; ++i)
{
    double yp = a * pow(vecX[i], 2) + b * vecX[i] + c;
    double err = abs(vecY[i] - yp) / vecY[i];
    cout << setw(6) << i;
    cout << setw(12) << vecY[i];
    cout << setw(12) << yp;
    cout << setw(10) << err * 100 << " %";
    cout << endl;
}
```

$$y = ax^2 + bx + c$$

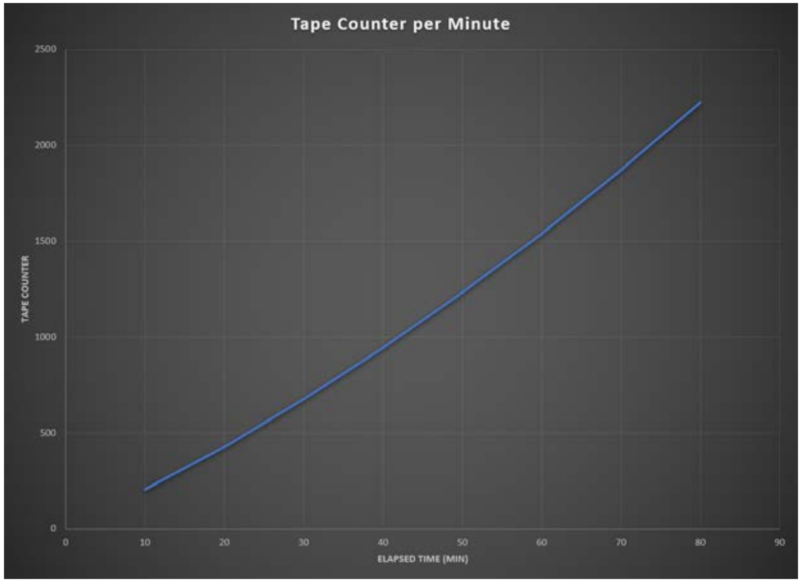
yp = y *predicted*
"the estimated y value"

Run Lab 5 - Method of Least Squares



$$y = 10.4643x^2 + 194.2262x + 0.1429$$

Elapsed Time (10 mins each)	Observed Counter	Estimated Counter
1	205	204.8333
2	430	430.4524
3	677	677.0000
4	945	944.4762
5	1233	1232.8810
6	1542	1542.2143
7	1872	1872.4762
8	2224	2223.6667



Verify Lab 5 - Method of Least Squares

via Difference Tables

$$y = 10.5833x^2 + 193.1786x + 1.8214$$

via Least Squares

$$y = 10.4643x^2 + 194.2262x + 0.1429$$

x	y Obs	y Est DT	% Err	y Est LS	% Err
1	205	205.5833	-0.2846%	204.8334	0.0813%
2	430	430.5119	-0.1190%	430.4525	-0.1052%
3	677	676.6071	0.0580%	677.0002	0.0000%
4	945	943.8690	0.1197%	944.4765	0.0554%
5	1233	1,232.2976	0.0570%	1232.8814	0.0096%
6	1542	1,541.8929	0.0069%	1542.2149	-0.0139%
7	1872	1,872.6548	-0.0350%	1872.477	-0.0255%
8	2224	2,224.5833	-0.0262%	2223.6677	0.0149%
		mean:	-0.0279%		0.0021%

10x better!

The **method of least squares** guarantees the <u>best</u> possible fit between the observed data and the polynomial form chosen to model the data

Now you know…

- Model fitting starts by assuming the degree of an appropriate curve that reasonably matches the data
- How to create **difference tables** for constant, linear, quadratic, cubic, and quartic terms along with their expected average
- Gauss's **Method of Least Squares** refers to shaping the approximating curve to ***minimize the total deviations*** between the observed data and points on that curve
  - The function to minimize must be expanded, and then **partial derivatives** must be found for each coefficient of the curve
  - Matrix algebra (Cramer's Rule) can then solve for the coefficients
- The **% relative error** measures the "goodness of fit" of a model to experimental observations