Essential Math for Data Analysis Using Excel Online

Module 1, Lab 2: Sigma Notation

Learning Objectives

- Practice sigma notation with Excel Online.

Description

Learners will work with a set of numbers and practice sigma notation. They will practice sigma(x), sigma($x – 1$), sigma($x^2 – 1$).

Data set

Mod1Lab1.csv

Overview

Sigma notation is the basic building block for creating a lot of different analytics formulas. In this lab, we'll practice using our new favorite Greek letter to build simple formulas by hand and with Excel.

What You'll Need

To complete the lab, you will need the online version of Microsoft Excel.

Exercise 1: Simple Sums

1. Open the data set in Excel. Here's what the data look like:

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N

2. Find the sum, where x = age:

$$\sum (x)$$

3. For starters, mentally translate what this actually means.

- Sigma means "sum," so the formula is telling you to add up a bunch of stuff.
- x is the age variable, so it represents all the values in the age column.

4. Now do the actual math. If you wanted to do this one by hand, you'd simply add up the ages of Person 1, Person 2, etc., all the way to Person 20.

45 + 50 + 46 + 51 + 35 + 24 + 30 + 37 + 45 + 55 + 19 + 24 + 45 + 18 + 32 + 35 + 32 + 33 + 25 + 53

That's pretty unwieldy, and even if you punched all the values into a calculator, there are plenty of opportunities to make a mistake. That's why using a formula in Excel is a better bet.

5. Build a formula in Excel. First, add a new row for the sum of all the ages.

	A	B	C
			fx age sum
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N
22	age sum		

Add a sum formula to cell B22. Click into that cell, then enter the SUM function. The syntax works like this: **=SUM(first cell:last cell)**. Since you're summing all the values from the age column, the first cell is B2 (that's the age of Person 1) and the last cell is B21 (the age of Person 20). Type in =SUM(B2:B21).

Note: if you don't want to *type* out "B2:B21" then you can also type "=SUM()", place your cursor in between the parentheses, and highlight those cells with your mouse.

fx =SUM(B2:B21)

Hit Enter, and now cell B22 will show the sum $\sum(x)$.

f_x =SUM(B2:B21)

	A	B	C
1	id	age	entrylevel
2	1	45	N
3	2	50	N
4	3	46	Y
5	4	51	N
6	5	35	N
7	6	24	N
8	7	30	N
9	8	37	N
10	9	45	N
11	10	55	N
12	11	19	Y
13	12	24	Y
14	13	45	N
15	14	18	N
16	15	32	N
17	16	35	N
18	17	32	N
19	18	33	N
20	19	25	N
21	20	53	N
22	age sum	734	

Now that you're familiar with the basic syntax, the other exercises should go a bit quicker.

Exercise 2: Subtracting a Constant from Each Value

Let's say you want to subtract 1 from each person's age in the data set, and then add the new ages together.

1. Build your sigma notation formula. Since x is the age variable and you want to lower each age by 1 year, the inside of your formula will be $x - 1$. You're still using all the ages in the list for your sum, so the formula will still run from Person 1 up to Person 20. Here's the formula:

$$\sum(x - 1)$$

2. Mentally translate the formula. You're taking each age value, subtracting 1 from that value, and *then* summing all those new values. The first few terms of the sum would look like this:

$(45 - 1) + (50 - 1) + (46 - 1) + (51 - 1) + (35 - 1) +...$

And so on. Who has time to crunch all those numbers by hand? It's Excel formulas to the rescue again.

3. Create a new column in the data set for $x - 1$.

	f_x	age - 1		
	A	**B**	**C**	**D**
1	id	age	entrylevel	age - 1
2	1	45 N		
3	2	50 N		
4	3	46 Y		
5	4	51 N		
6	5	35 N		
7	6	24 N		
8	7	30 N		
9	8	37 N		
10	9	45 N		
11	10	55 N		
12	11	19 Y		
13	12	24 Y		
14	13	45 N		
15	14	18 N		
16	15	32 N		
17	16	35 N		
18	17	32 N		
19	18	33 N		
20	19	25 N		
21	20	53 N		
22	age sum	734		

4. In cell D2, use a formula to subtract 1 from the age of Person 1. To do this, click into cell D2 and type in **=B2-1**, since you're subtracting 1 from cell B2 (which is Person 1's age). Then hit Enter.

f_x =B2-1

	A	B	C	D
1	id	age	entrylevel	age - 1
2	1	45	N	44
3	2	50	N	
4	3	46	Y	
5	4	51	N	
6	5	35	N	
7	6	24	N	
8	7	30	N	
9	8	37	N	
10	9	45	N	
11	10	55	N	
12	11	19	Y	
13	12	24	Y	
14	13	45	N	
15	14	18	N	
16	15	32	N	
17	16	35	N	
18	17	32	N	
19	18	33	N	
20	19	25	N	
21	20	53	N	
22	age sum	734		

5. Thankfully, you don't need to do this a bunch of times. Instead, you can use the "fill series" trick to automatically populate every other cell in the "age – 1" column. Click on the tiny green square in the bottom-right corner of cell D2.

age - 1

44
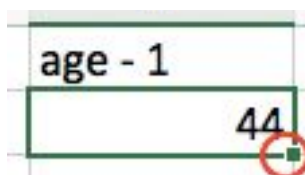
Now drag that square from Person 1 down to Person 20, and it will automatically populate each cell with the value $x – 1$. The program is smart enough to know that you mean to use this formula for each individual cell in column B, so it'll populate D3 with =B3-1, D4 with =B4-1, and so on.

	A	B	C	D
1	id	age	entrylevel	age - 1
2	1	45 N		44
3	2	50 N		49
4	3	46 Y		45
5	4	51 N		50
6	5	35 N		34
7	6	24 N		23
8	7	30 N		29
9	8	37 N		36
10	9	45 N		44
11	10	55 N		54
12	11	19 Y		18
13	12	24 Y		23
14	13	45 N		44
15	14	18 N		17
16	15	32 N		31
17	16	35 N		34
18	17	32 N		31
19	18	33 N		32
20	19	25 N		24
21	20	53 N		52
22	age sum	734		

6. The last step is to find the sum of these new values. In the same row you used for the sum earlier in Exercise 1, use the SUM function again to add all the values from column D. Type this into cell D22. (Once again, if you'd prefer, you can just click inside the parentheses in the SUM formula and highlight all the values in column D instead.)

f_x =SUM(D2:D21)

Hit Enter, and boom, you have your sum. It's 714 this time.

f_x =SUM(D2:D21)

	A	B	C	D
1	id	age	entrylevel	age - 1
2	1	45	N	44
3	2	50	N	49
4	3	46	Y	45
5	4	51	N	50
6	5	35	N	34
7	6	24	N	23
8	7	30	N	29
9	8	37	N	36
10	9	45	N	44
11	10	55	N	54
12	11	19	Y	18
13	12	24	Y	23
14	13	45	N	44
15	14	18	N	17
16	15	32	N	31
17	16	35	N	34
18	17	32	N	31
19	18	33	N	32
20	19	25	N	24
21	20	53	N	52
22	age sum	734		714

Exercise 3: Squaring a Value and Subtracting a Constant

Let's do one more, this time with a squared value. For whatever reason, let's say we want to square each person's age, subtract 1 from that squared value, and then sum all the new values.

1. Build your sigma notation formula. Since x is the age variable and you want to square each age and then lower the squared value by 1, the inside of your formula will be $x^2 - 1$. You're still using all the ages in the list for your sum, so the formula will still run from Person 1 down to Person 20. Here's the sigma formula:

$$\sum (x^2 - 1)$$

2. Just to make sure you know what's happening mathematically, mentally translate the formula again. The first few terms of the sum would look like this:

$(45^2 - 1) + (50^2 - 1) + (46^2 - 1) + (51^2 - 1) + (35^2 - 1) +...$

And so on.

3. Create a new column in the data set for $x^2 - 1$.

fx age squared - 1

	A	B	C	D	E
1	id	age	entrylevel	age - 1	age squared - 1
2	1	45	N	44	
3	2	50	N	49	
4	3	46	Y	45	
5	4	51	N	50	
6	5	35	N	34	
7	6	24	N	23	
8	7	30	N	29	
9	8	37	N	36	
10	9	45	N	44	
11	10	55	N	54	
12	11	19	Y	18	
13	12	24	Y	23	
14	13	45	N	44	
15	14	18	N	17	
16	15	32	N	31	
17	16	35	N	34	
18	17	32	N	31	
19	18	33	N	32	
20	19	25	N	24	
21	20	53	N	52	
22	age sum	734		714	
23					

4. Formula time. Click into cell E2 and type in **=B2^2-1**, since you're squaring cell B2 and subtracting 1.

f_x =B2^2-1

	A	B	C	D	E
1	id	age	entrylevel	age - 1	age squared - 1
2	1	45 N		44	2024
3	2	50 N		49	
4	3	46 Y		45	
5	4	51 N		50	
6	5	35 N		34	
7	6	24 N		23	
8	7	30 N		29	
9	8	37 N		36	
10	9	45 N		44	
11	10	55 N		54	
12	11	19 Y		18	
13	12	24 Y		23	
14	13	45 N		44	
15	14	18 N		17	
16	15	32 N		31	
17	16	35 N		34	
18	17	32 N		31	
19	18	33 N		32	
20	19	25 N		24	
21	20	53 N		52	
22	age sum	734		714	

5. Use the "fill series" trick again to automatically populate every other cell in the "age squared – 1" column. Click on the tiny green square in the bottom-right corner of cell E2 and drag it all the way down to E21.

f_x =B2^2-1

	A	B	C	D	E
1	id	age	entrylevel	age - 1	age squared - 1
2	1	45 N		44	2024
3	2	50 N		49	2499
4	3	46 Y		45	2115
5	4	51 N		50	2600
6	5	35 N		34	1224
7	6	24 N		23	575
8	7	30 N		29	899
9	8	37 N		36	1368
10	9	45 N		44	2024
11	10	55 N		54	3024
12	11	19 Y		18	360
13	12	24 Y		23	575
14	13	45 N		44	2024
15	14	18 N		17	323
16	15	32 N		31	1023
17	16	35 N		34	1224
18	17	32 N		31	1023
19	18	33 N		32	1088
20	19	25 N		24	624
21	20	53 N		52	2808
22	age sum	734		714	

6. In cell E22, use the SUM function to add up all the values from column E.

f_x =SUM(E2:E21)

Hit Enter, and you've got your sum.

f_x =SUM(E2:E21)

	A	B	C	D	E
1	id	age	entrylevel	age - 1	age squared - 1
2	1	45	N	44	2024
3	2	50	N	49	2499
4	3	46	Y	45	2115
5	4	51	N	50	2600
6	5	35	N	34	1224
7	6	24	N	23	575
8	7	30	N	29	899
9	8	37	N	36	1368
10	9	45	N	44	2024
11	10	55	N	54	3024
12	11	19	Y	18	360
13	12	24	Y	23	575
14	13	45	N	44	2024
15	14	18	N	17	323
16	15	32	N	31	1023
17	16	35	N	34	1224
18	17	32	N	31	1023
19	18	33	N	32	1088
20	19	25	N	24	624
21	20	53	N	52	2808
22	age sum	734		714	29424