SDSC 2102
STATISTICAL METHODS AND DATA ANALYSIS

Lab 1, 2 – Data visualization and analysis

Aims

- A. Learn descriptive statistics
  - Calculate mean, median, mode;
  - Calculate maximum, minimum, range, sample variance, sample standard deviation.
- B. Learn data representation
  - Plot the histogram;
  - Plot the box-whisker.

Part A. Descriptive Statistics

- A.1 Measures of Central Tendency
  - Mean: the sum of the data divided by the number of data points in the data set, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;
  - Median: the middle number when the data set is arranged in ascending (or descending) order.
  - Mode: the most frequently occurring number in the data set.

  - Example:1, 2, 2, 3, 5, 7, 9

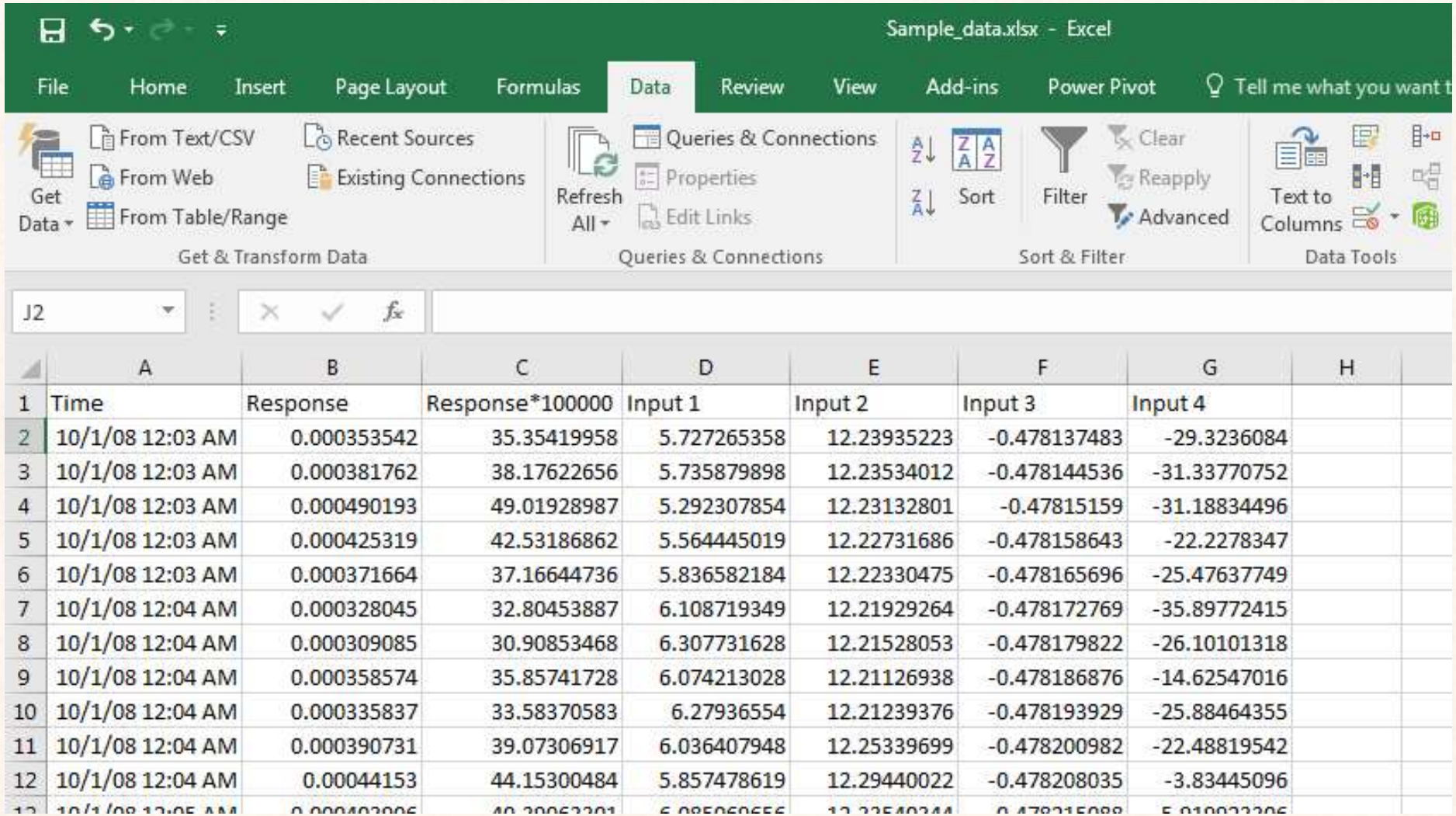
    mean: (1+2+2+3+5+7+9)/7 = 4.14285

    median: 3

    mode: 2

Part A.1 Mean

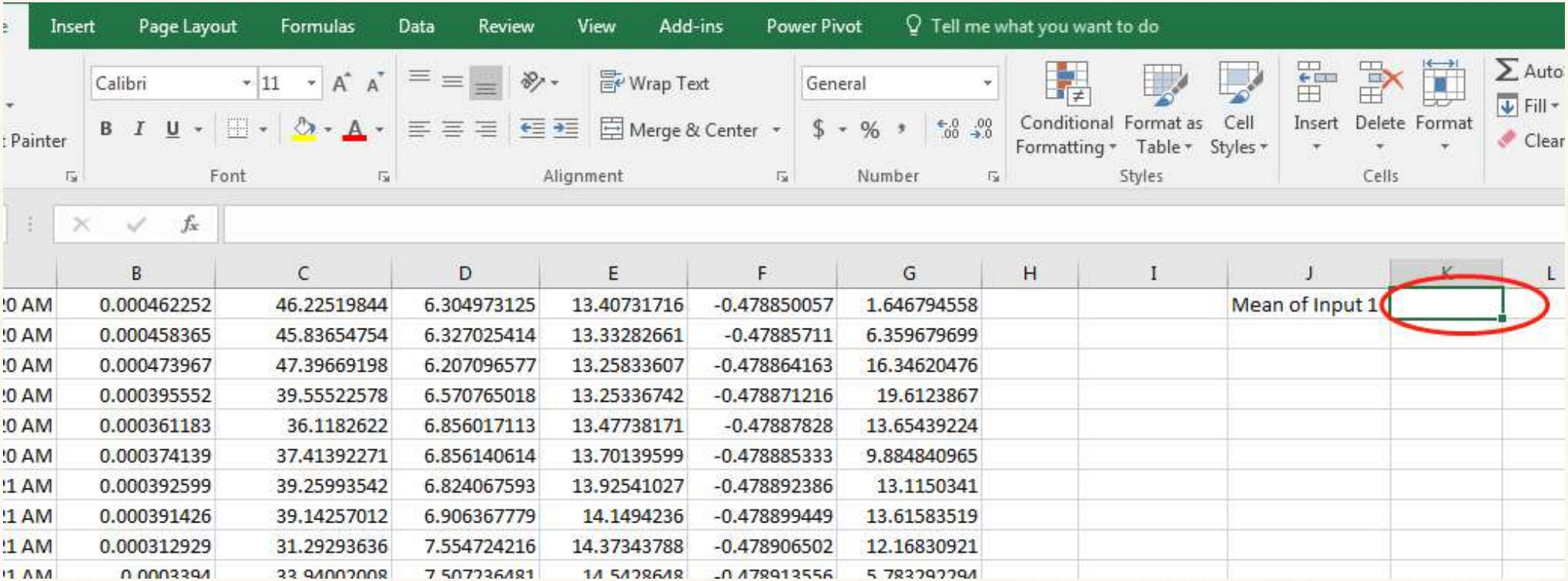
- Start excel and open the file Sample_data.xlsx



	A	B	C	D	E	F	G	H
1	Time	Response	Response*100000	Input 1	Input 2	Input 3	Input 4	
2	10/1/08 12:03 AM	0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084	
3	10/1/08 12:03 AM	0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752	
4	10/1/08 12:03 AM	0.000490193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496	
5	10/1/08 12:03 AM	0.000425319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347	
6	10/1/08 12:03 AM	0.000371664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749	
7	10/1/08 12:04 AM	0.000328045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415	
8	10/1/08 12:04 AM	0.000309085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318	
9	10/1/08 12:04 AM	0.000358574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016	
10	10/1/08 12:04 AM	0.000335837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355	
11	10/1/08 12:04 AM	0.000390731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542	
12	10/1/08 12:04 AM	0.00044153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096	

Part A.1 Mean

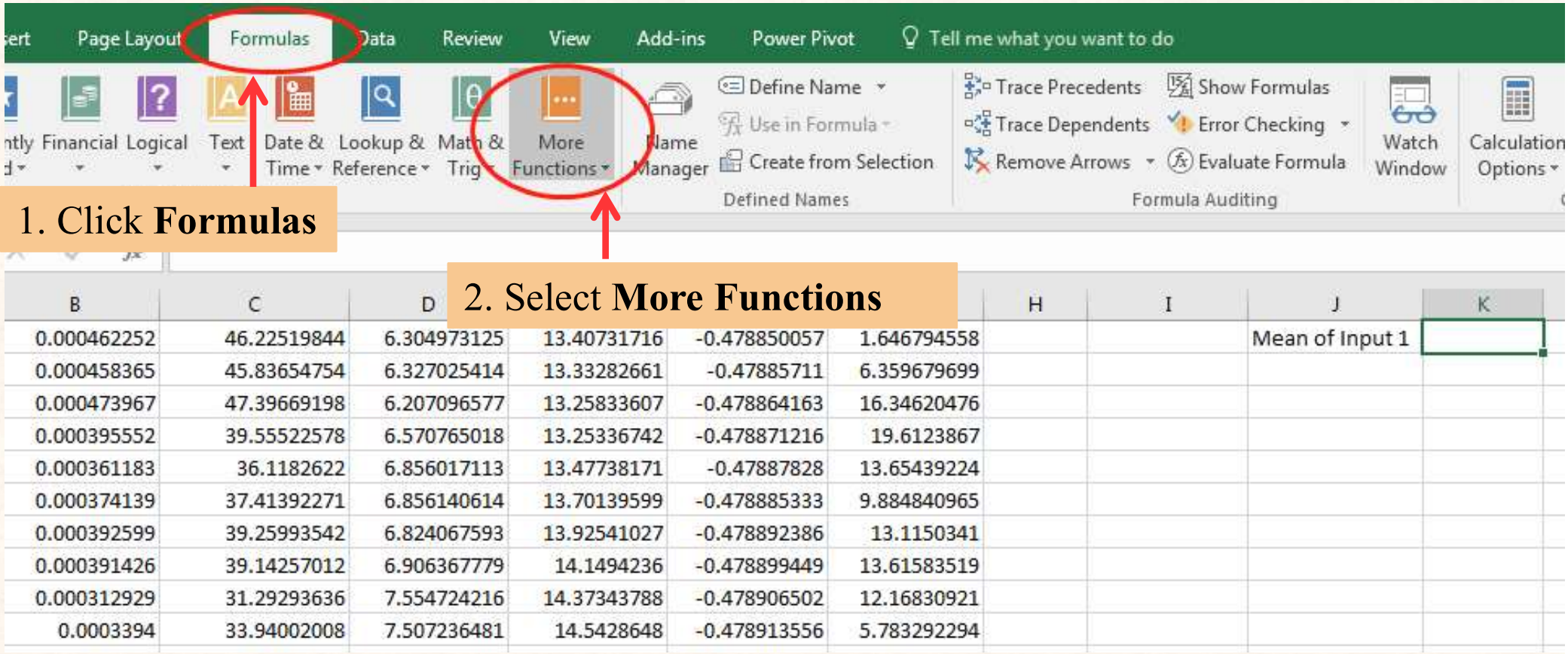
- Step 1: place your cursor in the cell labeled Mean



	B	C	D	E	F	G
0 AM	0.000462252	46.22519844	6.304973125	13.40731716	-0.478850057	1.646794558
0 AM	0.000458365	45.83654754	6.327025414	13.33282661	-0.47885711	6.359679699
0 AM	0.000473967	47.39669198	6.207096577	13.25833607	-0.478864163	16.34620476
0 AM	0.000395552	39.55522578	6.570765018	13.25336742	-0.478871216	19.6123867
0 AM	0.000361183	36.1182622	6.856017113	13.47738171	-0.47887828	13.65439224
0 AM	0.000374139	37.41392271	6.856140614	13.70139599	-0.478885333	9.884840965
1 AM	0.000392599	39.25993542	6.824067593	13.92541027	-0.478892386	13.1150341
1 AM	0.000391426	39.14257012	6.906367779	14.1494236	-0.478899449	13.61583519
1 AM	0.000312929	31.29293636	7.554724216	14.37343788	-0.478906502	12.16830921
1 AM	0.0003394	33.94002008	7.507236481	14.5428648	-0.478913556	5.783292294

Part A.1 Mean

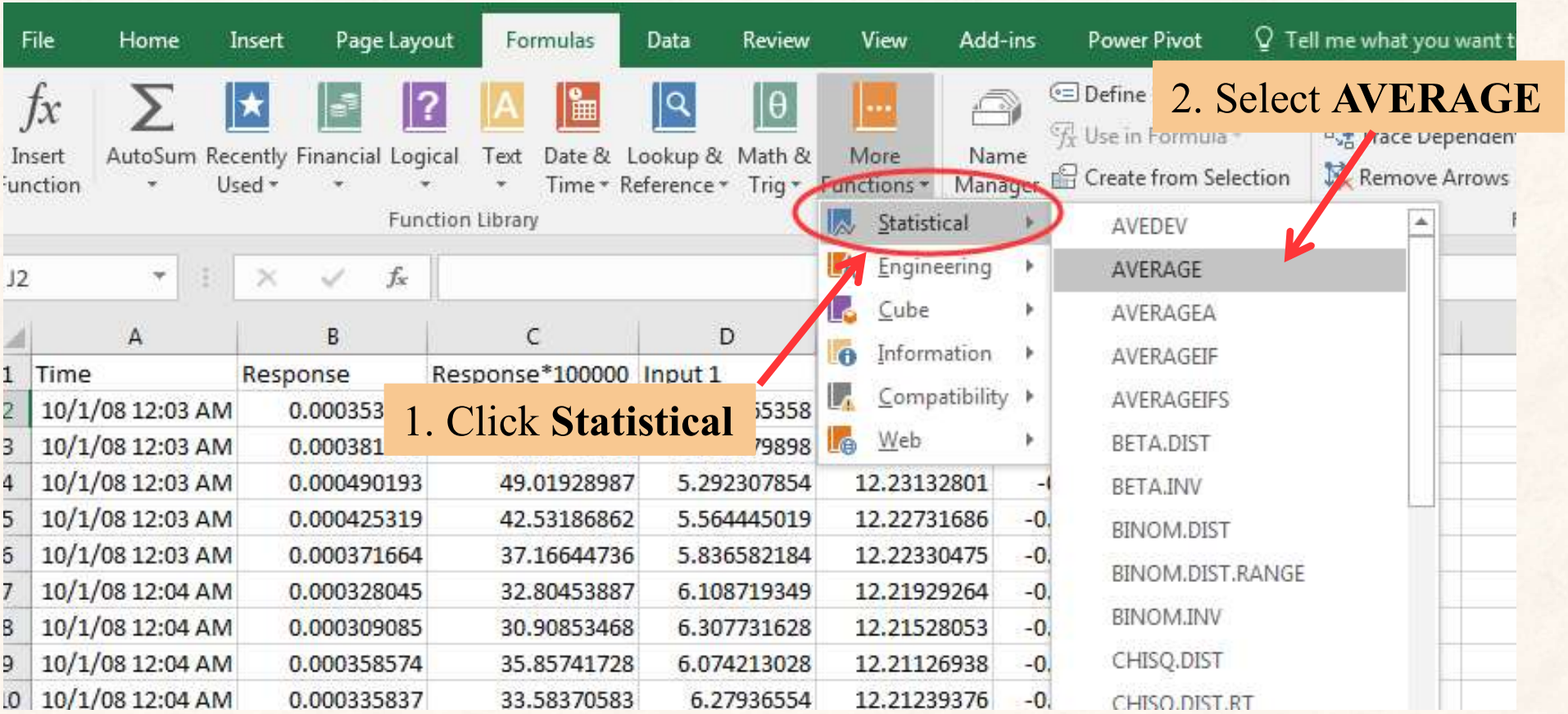
- Step 2: select More Functions … from the Formulas menu
  (or click on the paste function - *fx* - button on the menu)



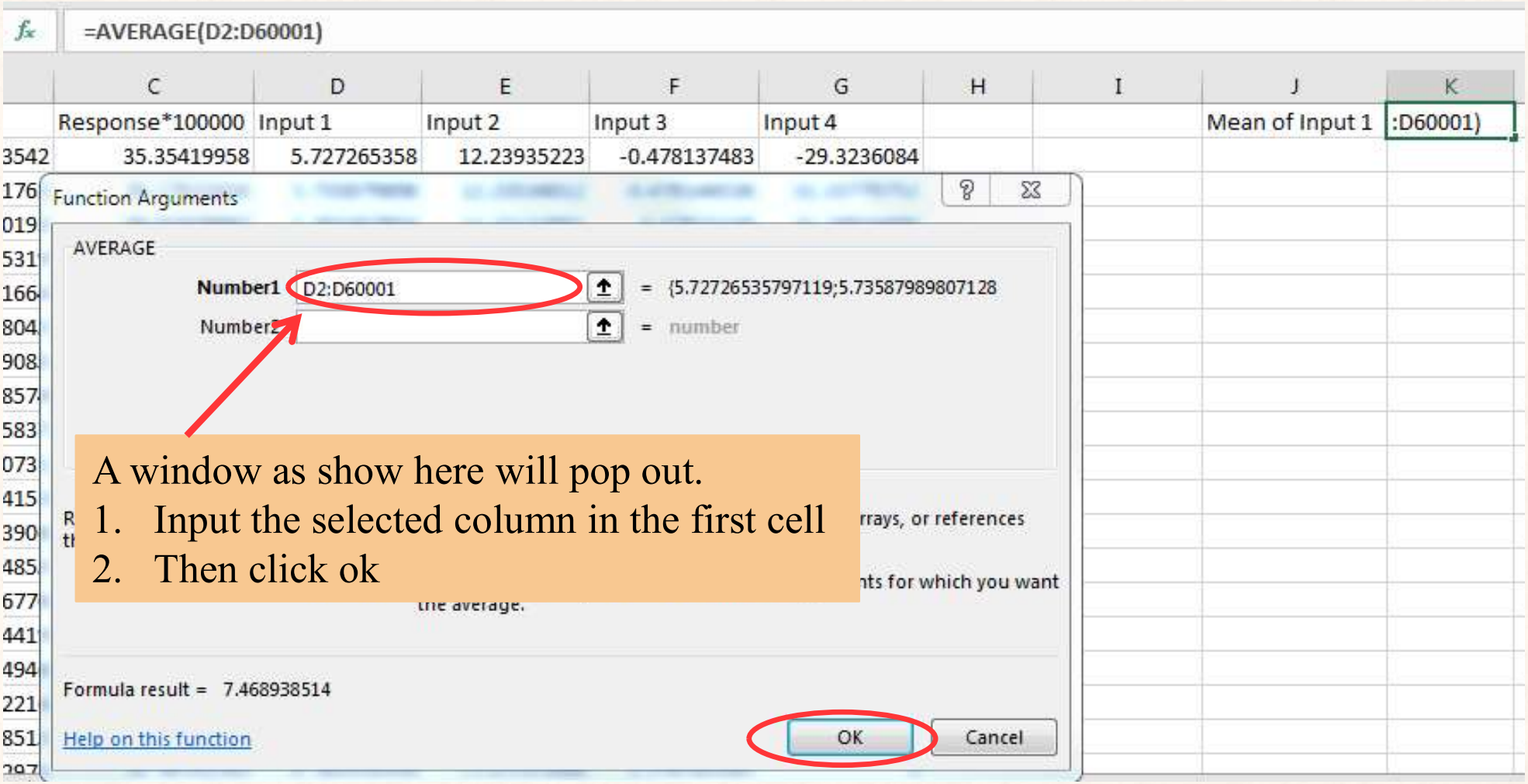
B	C	D				H	I	J	K
0.000462252	46.22519844	6.304973125	13.40731716	-0.478850057	1.646794558			Mean of Input 1	
0.000458365	45.83654754	6.327025414	13.33282661	-0.47885711	6.359679699				
0.000473967	47.39669198	6.207096577	13.25833607	-0.478864163	16.34620476				
0.000395552	39.55522578	6.570765018	13.25336742	-0.478871216	19.6123867				
0.000361183	36.1182622	6.856017113	13.47738171	-0.47887828	13.65439224				
0.000374139	37.41392271	6.856140614	13.70139599	-0.478885333	9.884840965				
0.000392599	39.25993542	6.824067593	13.92541027	-0.478892386	13.1150341				
0.000391426	39.14257012	6.906367779	14.1494236	-0.478899449	13.61583519				
0.000312929	31.29293636	7.554724216	14.37343788	-0.478906502	12.16830921				
0.0003394	33.94002008	7.507236481	14.5428648	-0.478913556	5.783292294				

Part A.1 Mean

- Step 3: select Function **AVERAGE** under the statistical menu



Part A.1 Mean



A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

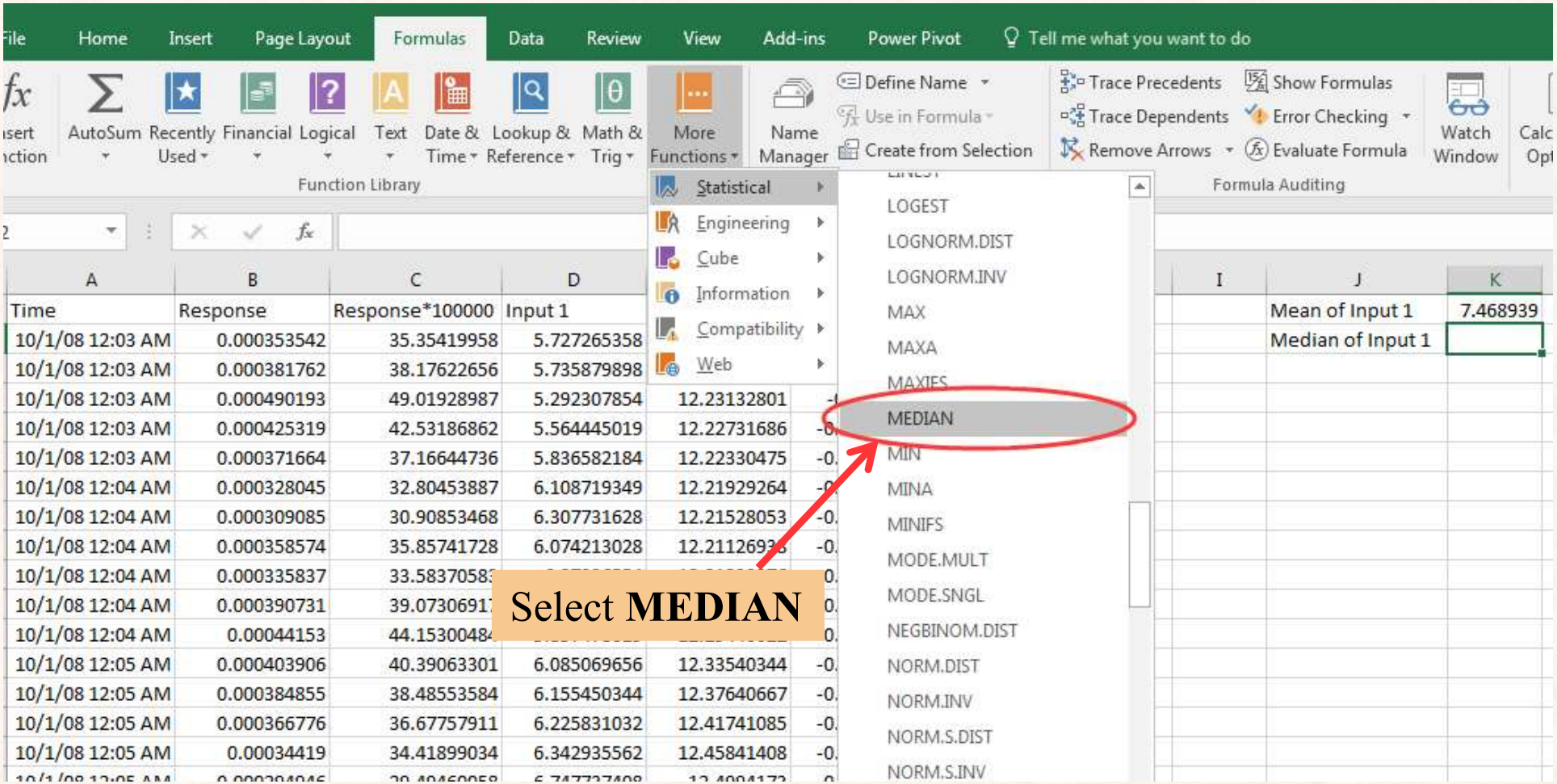
Part A.1 Mean

f_x =AVERAGE(D2:D60001)

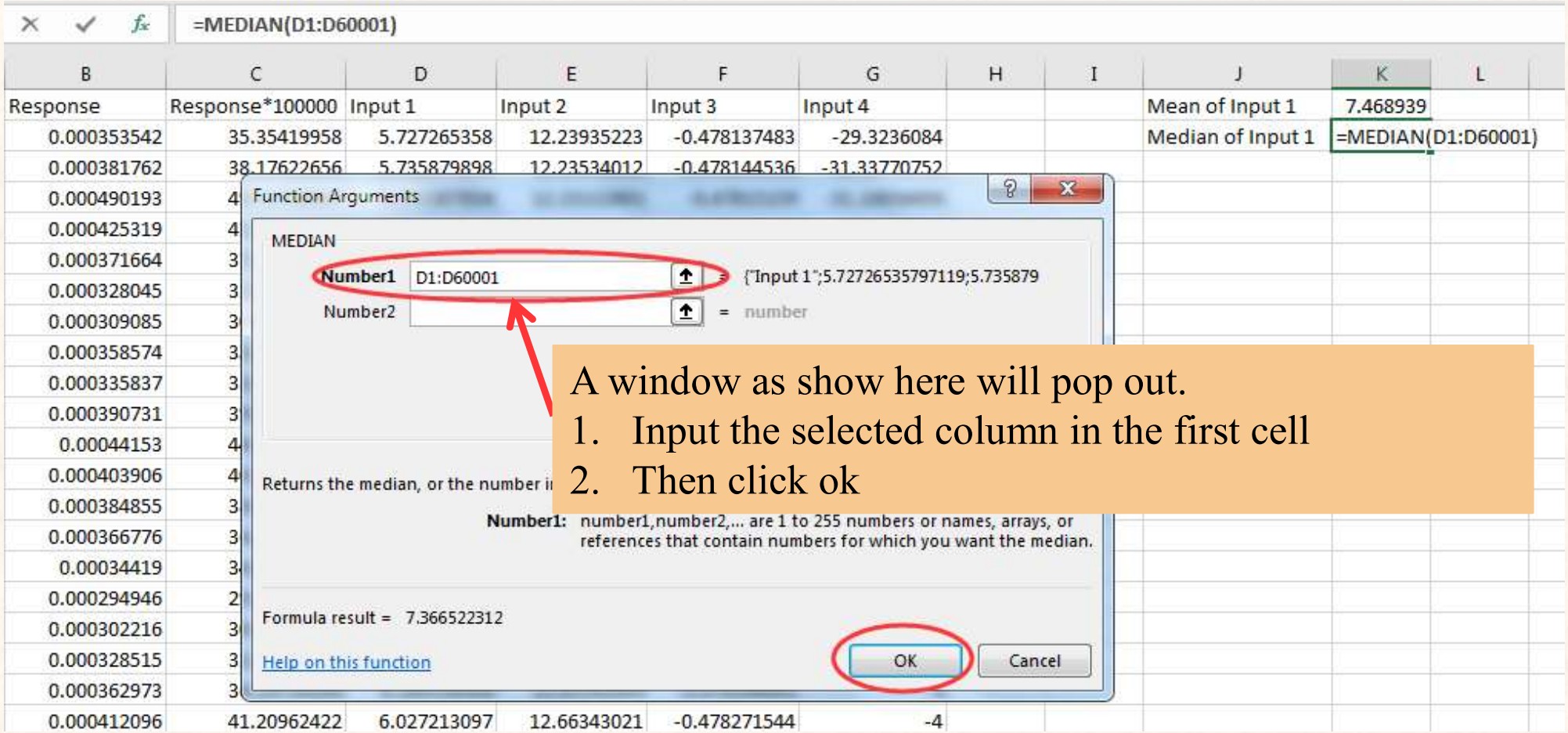
	C	D	E	F	G	H	I	J	K
	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.4689385
3542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084				
1762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752				
0193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496				
5319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
1664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
8045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
9085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
5837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
0731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
4153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
3906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
4855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				

Get the final result of Mean

Part A.1 Median



Part A.1 Median



A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

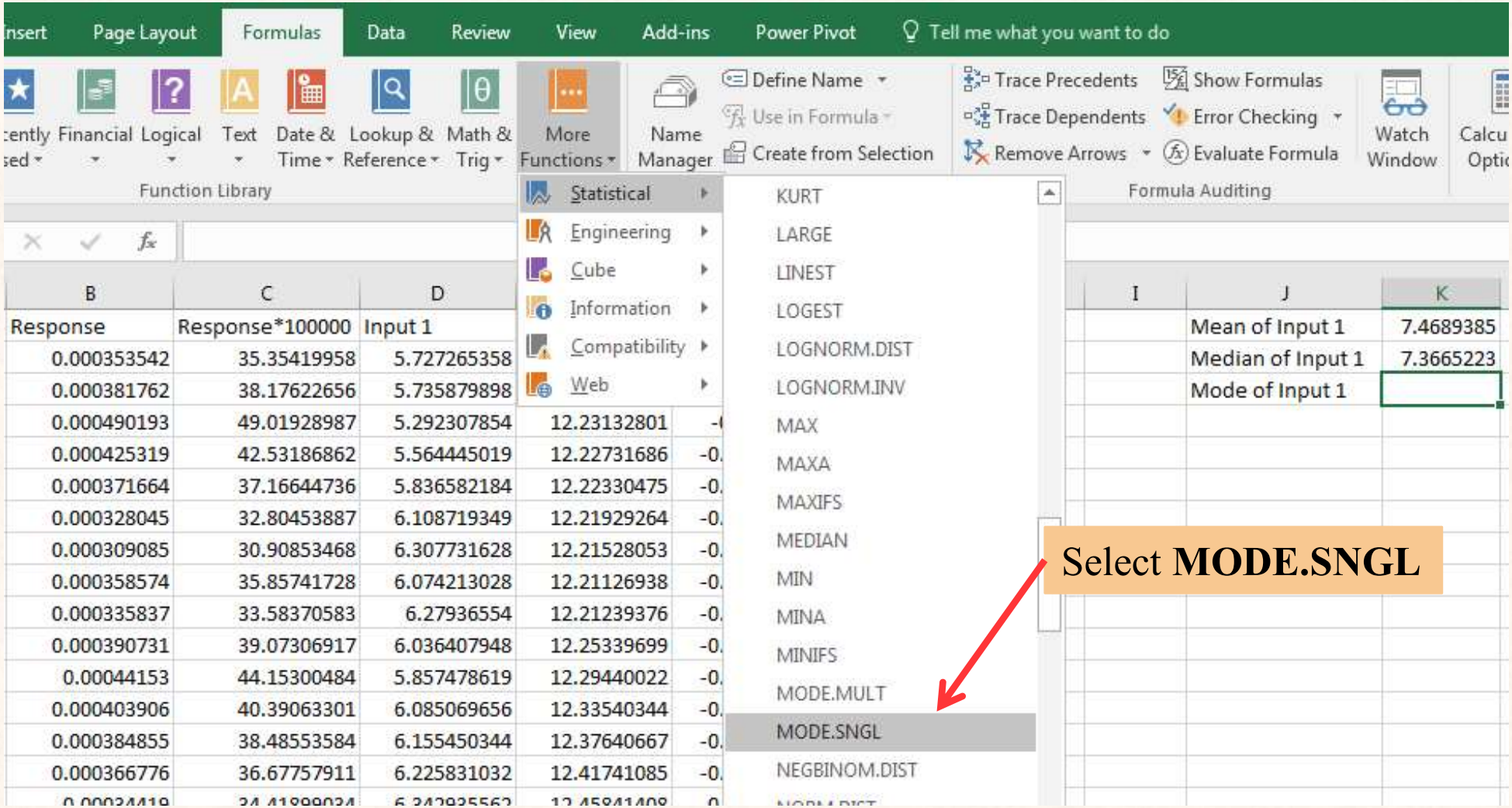
Part A.1 Median

=MEDIAN(D1:D60001)

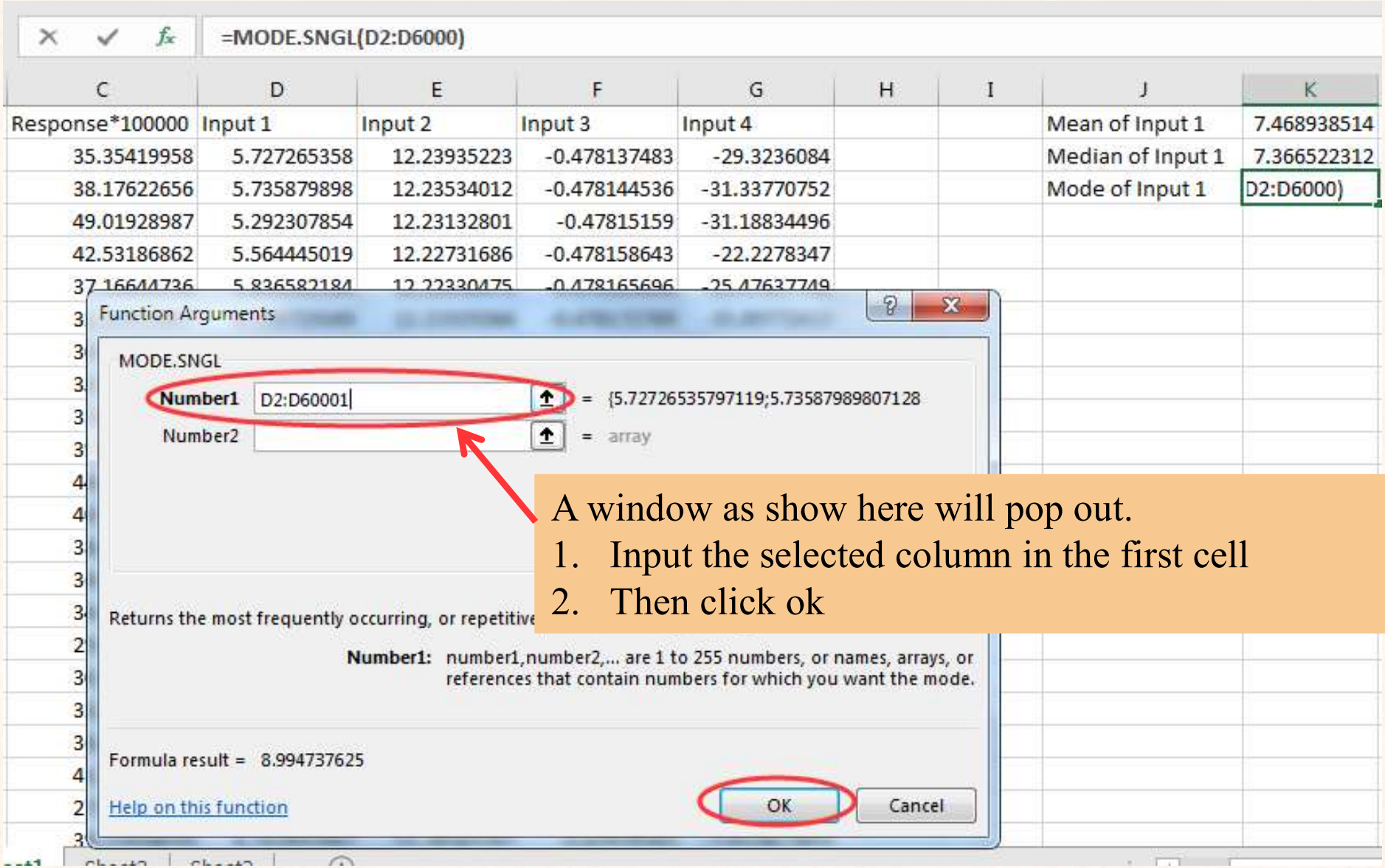
B	C	D	E	F	G	H	I	J	K
Response	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468939
0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752				
0.000490193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496				
0.000425319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
0.000371664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
0.000328045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
0.000309085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
0.000358574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
0.000335837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
0.000390731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
0.00044153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
0.000403906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
0.000384855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				
0.000366776	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763				

Get the final result of Median

Part A.1 Mode


Insert
Page Layout
Formulas
Data
Review
View
Add-ins
Power Pivot
Tell me what you want to do
Financial
Logical
Text
Date & Time
Lookup & Reference
Math & Trig
More Functions
Name Manager
Define Name
Use in Formula
Create from Selection
Trace Precedents
Trace Dependents
Remove Arrows
Show Formulas
Error Checking
Evaluate Formula
Watch Window
Function Library
Formula Auditing
Statistical
Engineering
Cube
Information
Compatibility
Web
KURT
LARGE
LINEST
LOGEST
LOGNORM.DIST
LOGNORM.INV
MAX
MAXA
MAXIFS
MEDIAN
MIN
MINA
MINIFS
MODE.MULT
MODE.SNGL
NEGBINOM.DIST
Response
Response*100000
Input 1
Mean of Input 1
7.4689385
Median of Input 1
7.3665223
Mode of Input 1
Select MODE.SNGL

Part A.1 Mode



A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.1 Mode

=MODE.SNGL(D2:D60001)

C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496				
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				
36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763				

Get the final result of Mode

Part A. Descriptive Statistics (cont'd)

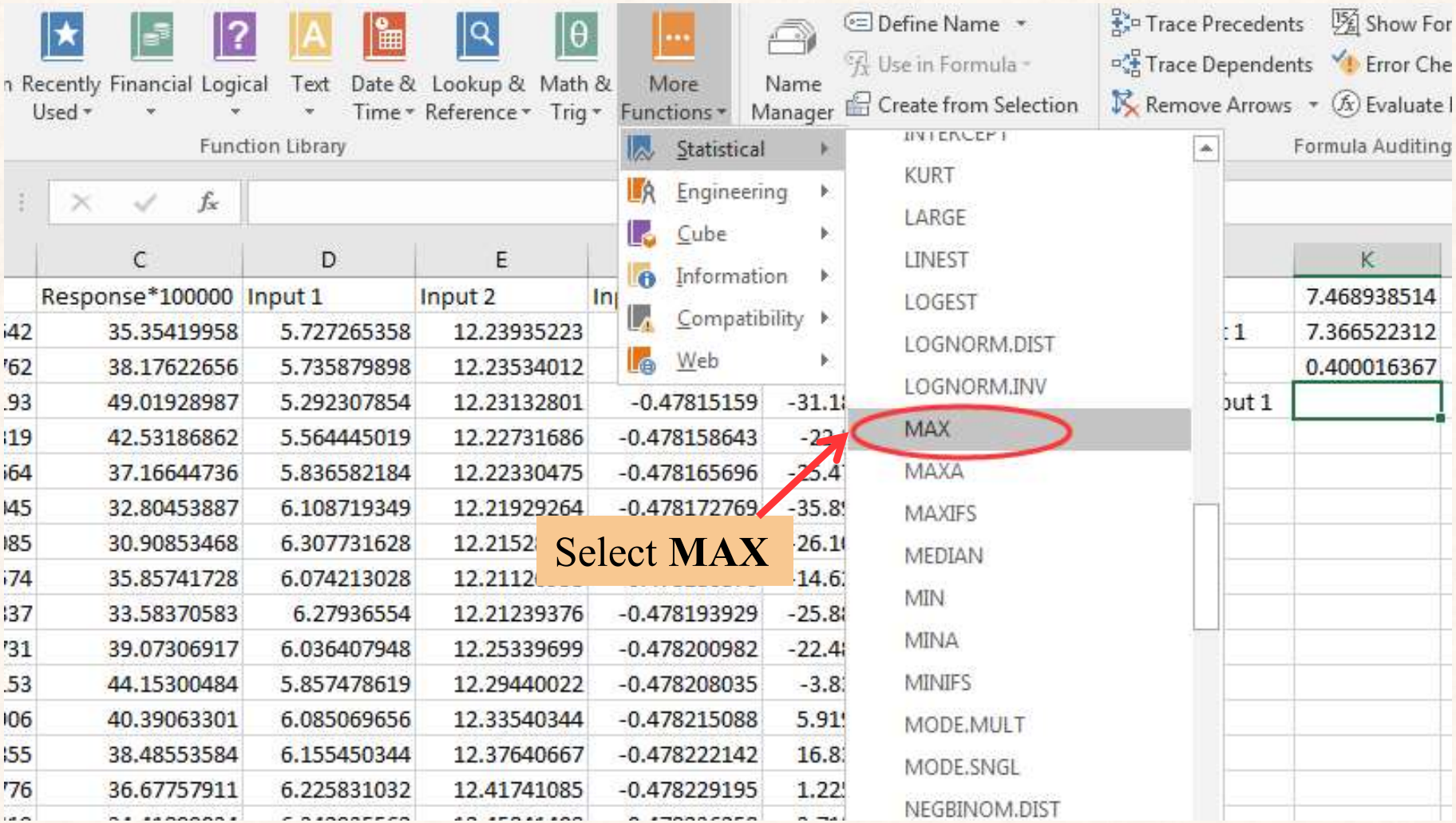
- A.2 Measures of Variability
  - Maximum and Minimum: the largest number and the smallest number in the data set
  - Range: maximum minus minimum
  - Sample Variance: the sum of each data points from the mean and divided by the number of data points minus one

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})$$

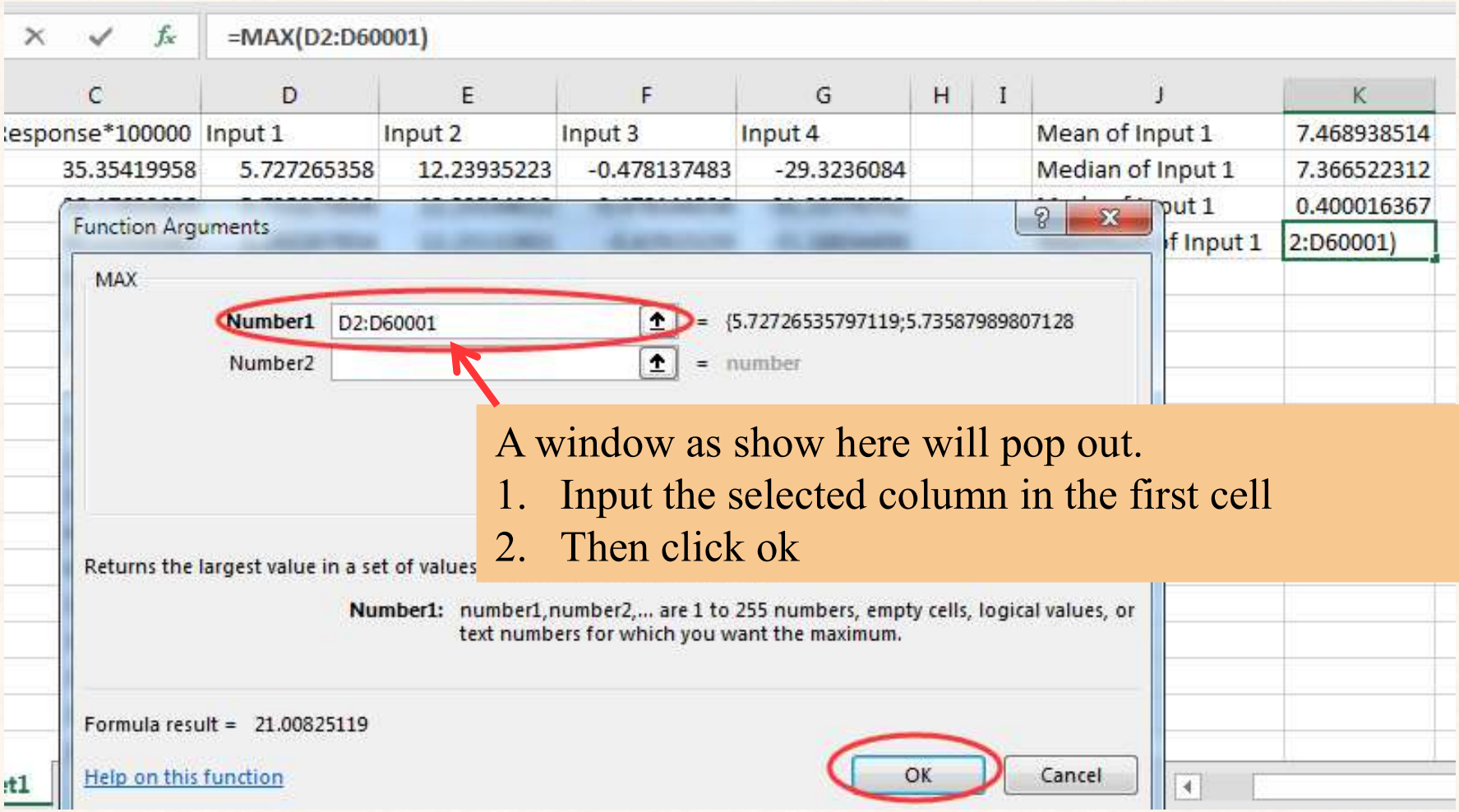
  - Sample Standard Deviation: the positive square root of the sample variance

$$\sigma = \sqrt{s^2}$$

Part A.2 Maximum & Minimum



Part A.2 Maximum & Minimum



A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

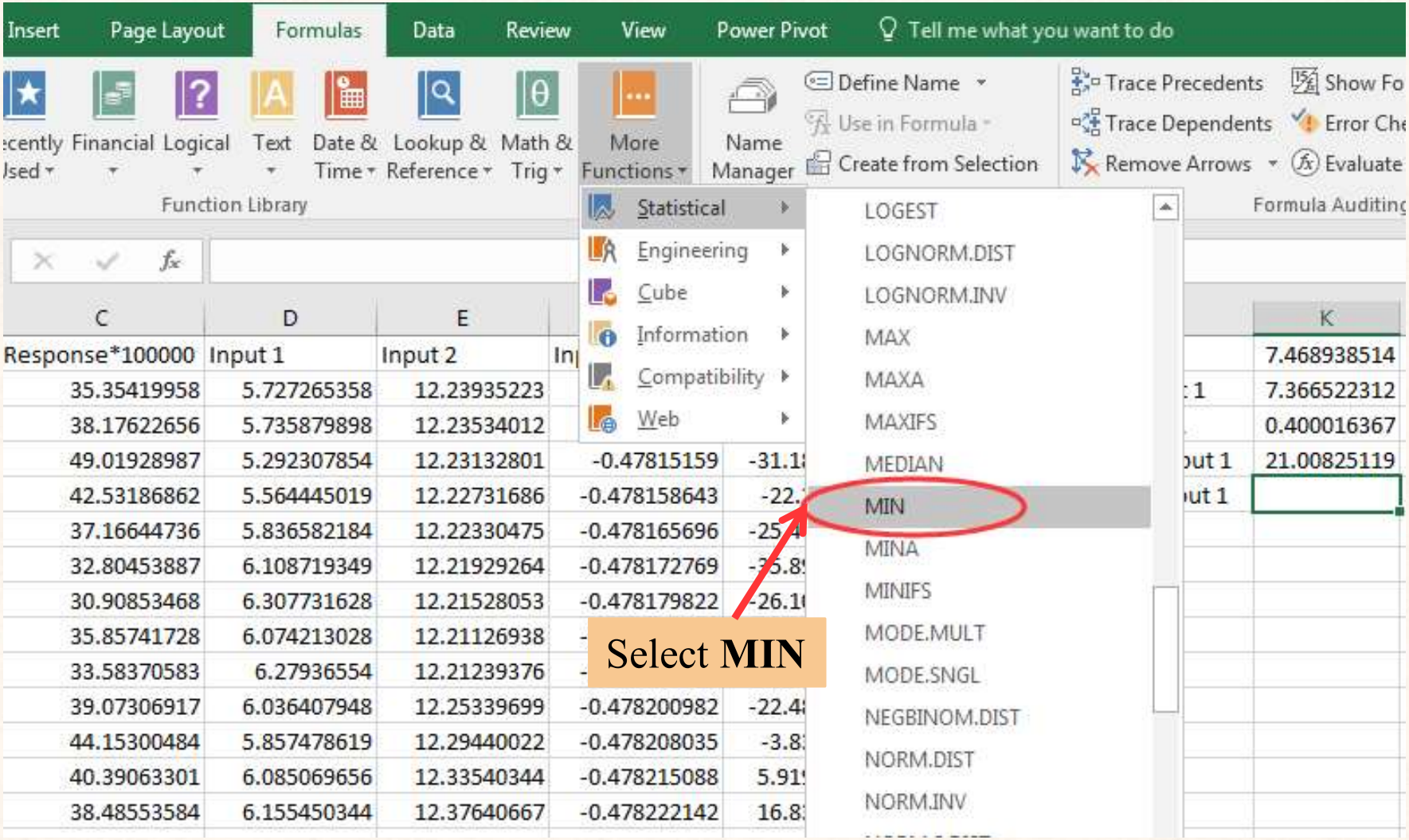
Part A.2 Maximum & Minimum

=MAX(D2:D60001)

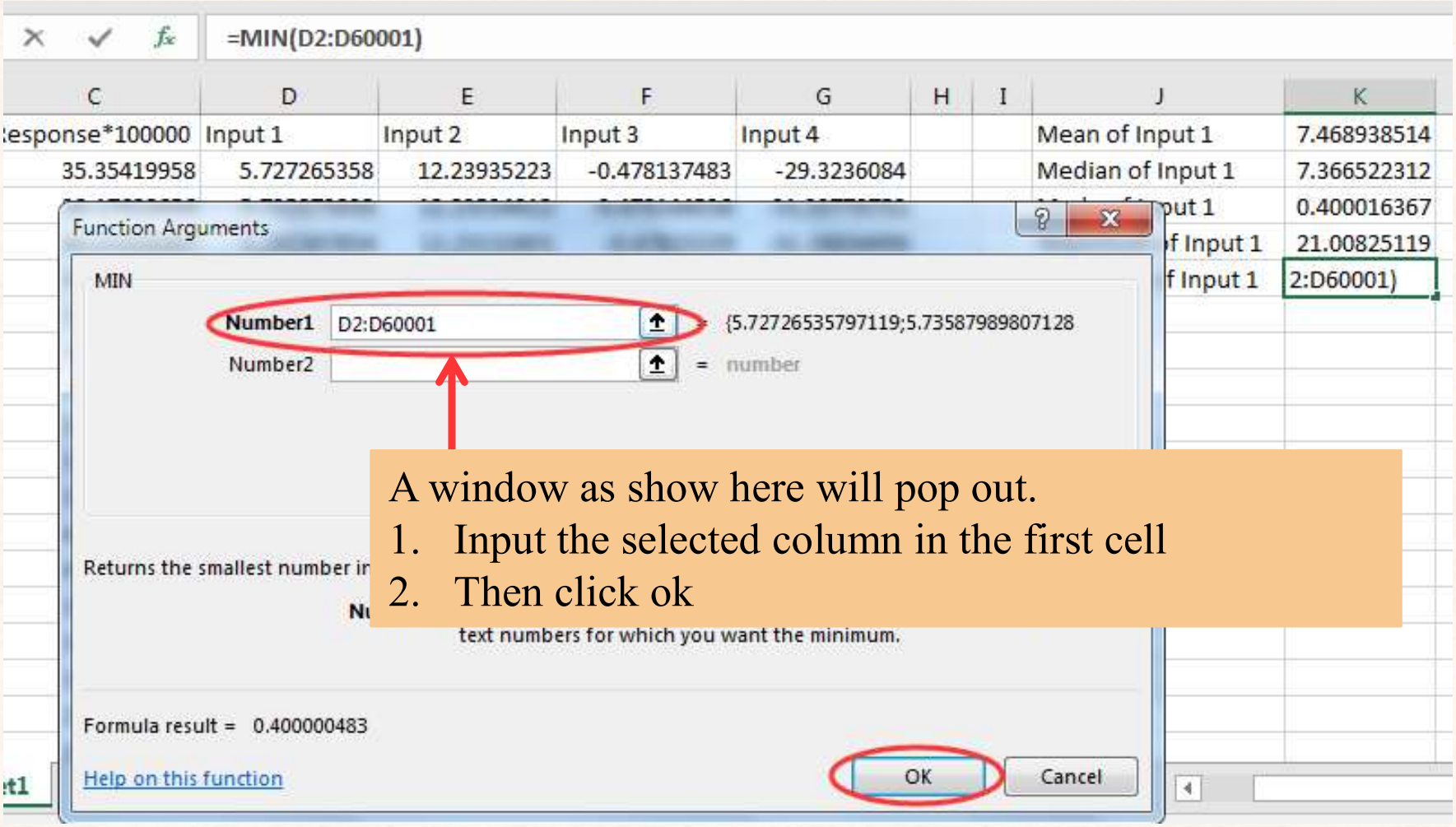
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				

Get the final result of Maximum

Part A.2 Maximum & Minimum



Part A.2 Maximum & Minimum



A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

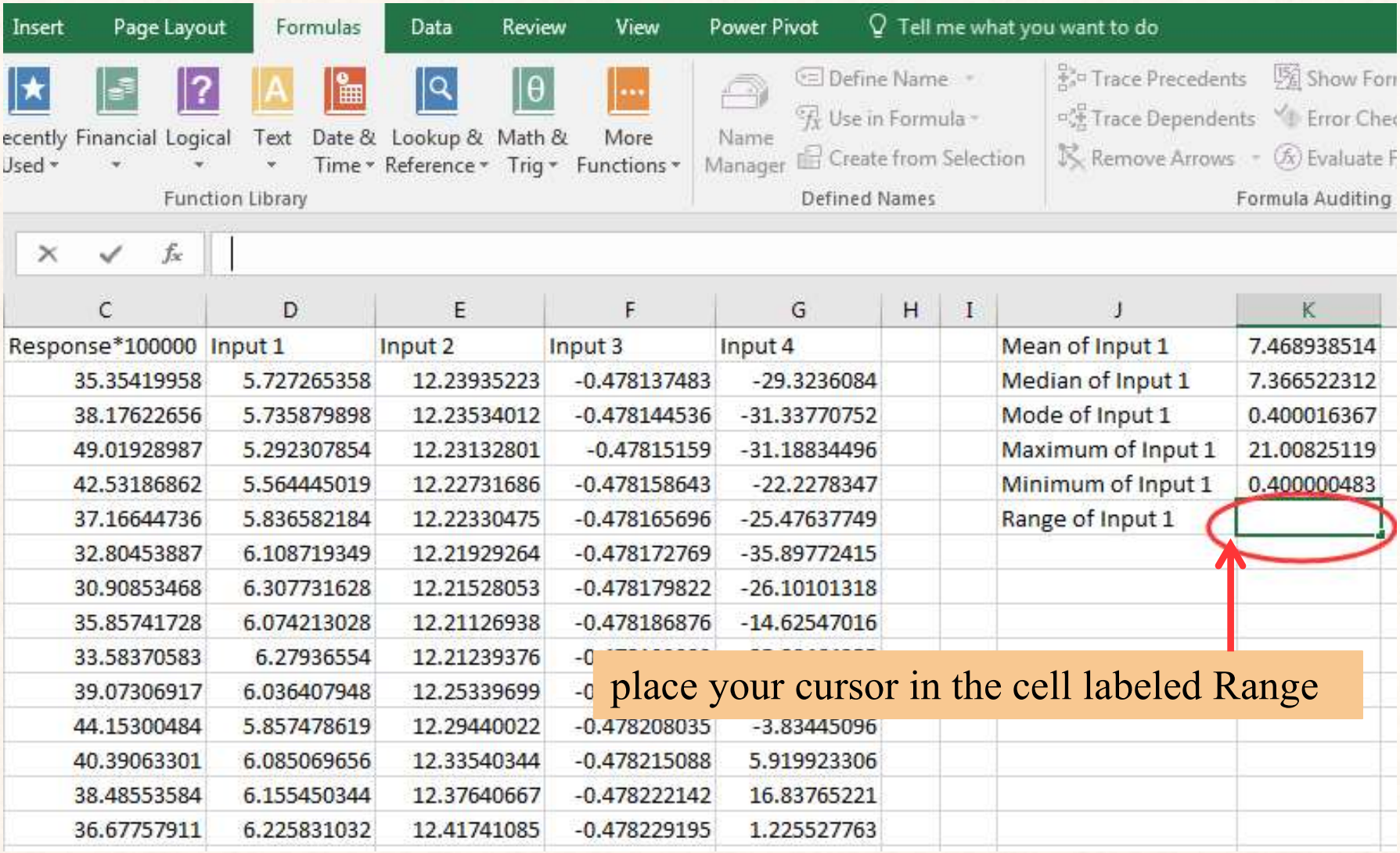
Part A.2 Maximum & Minimum

=MIN(D2:D60001)

C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				

Get the final result of Minimum

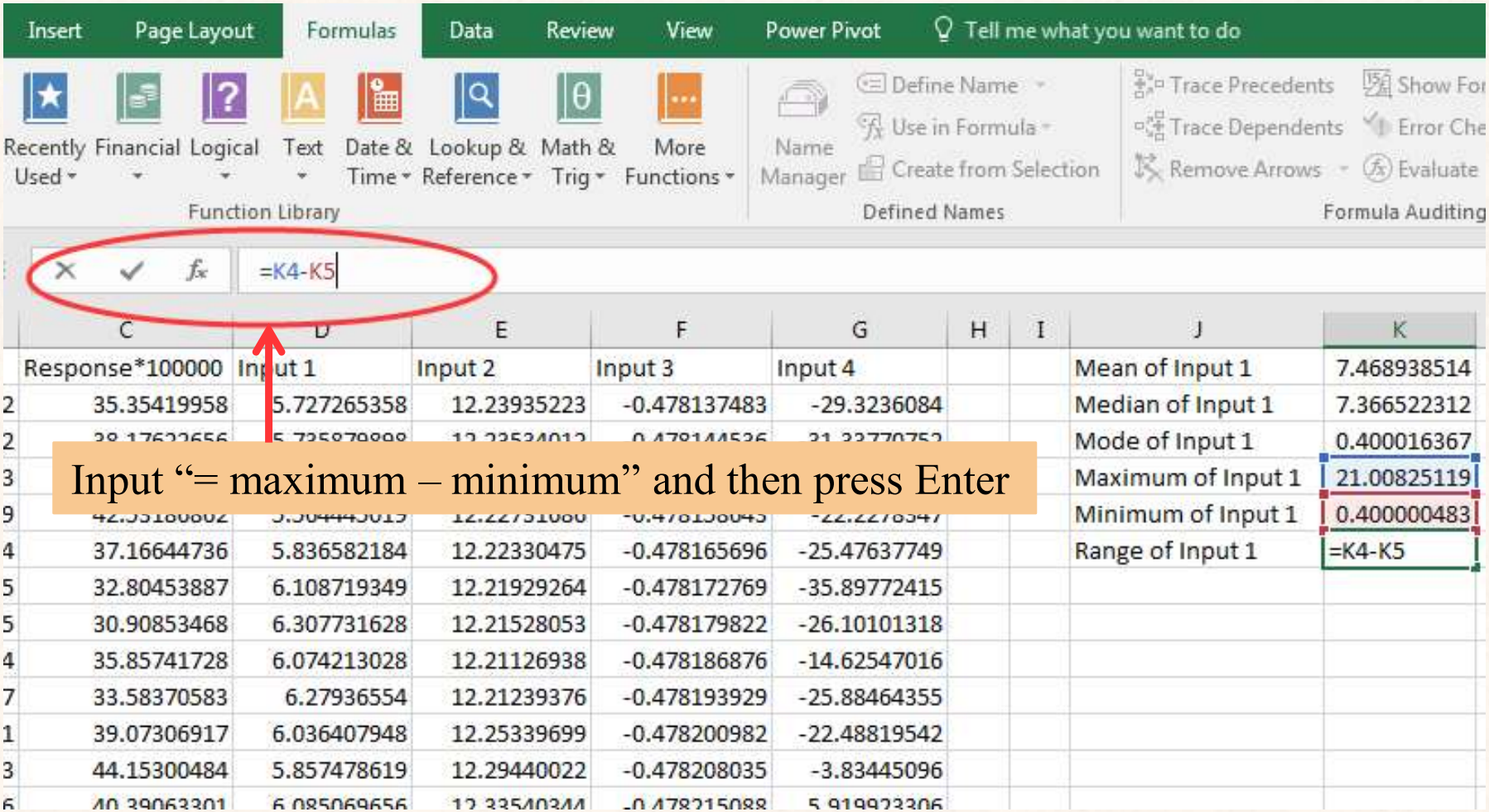
Part A.2 Range



C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0					
39.07306917	6.036407948	12.25339699	-0					
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				
36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763				

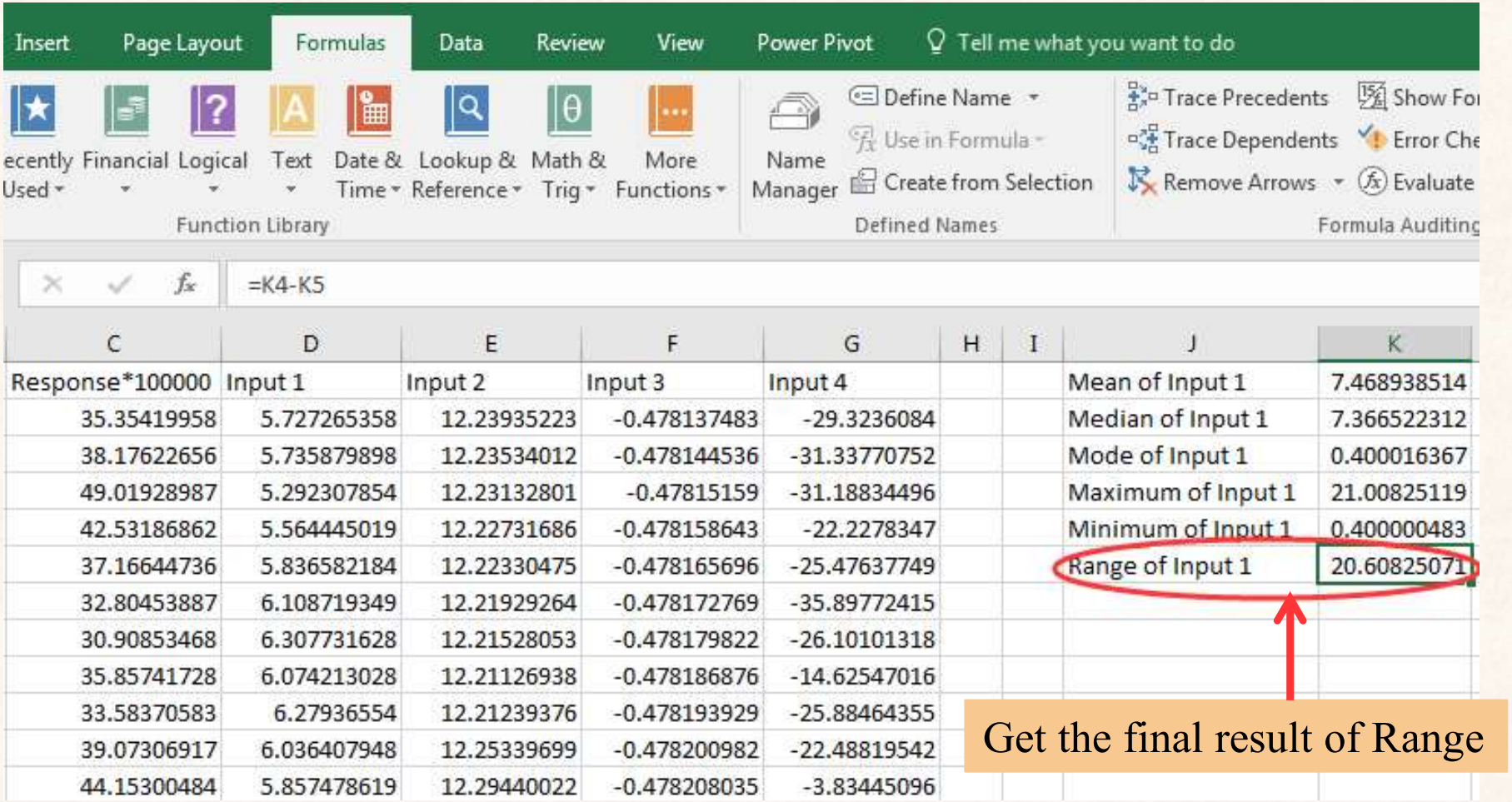
place your cursor in the cell labeled Range

Part A.2 Range



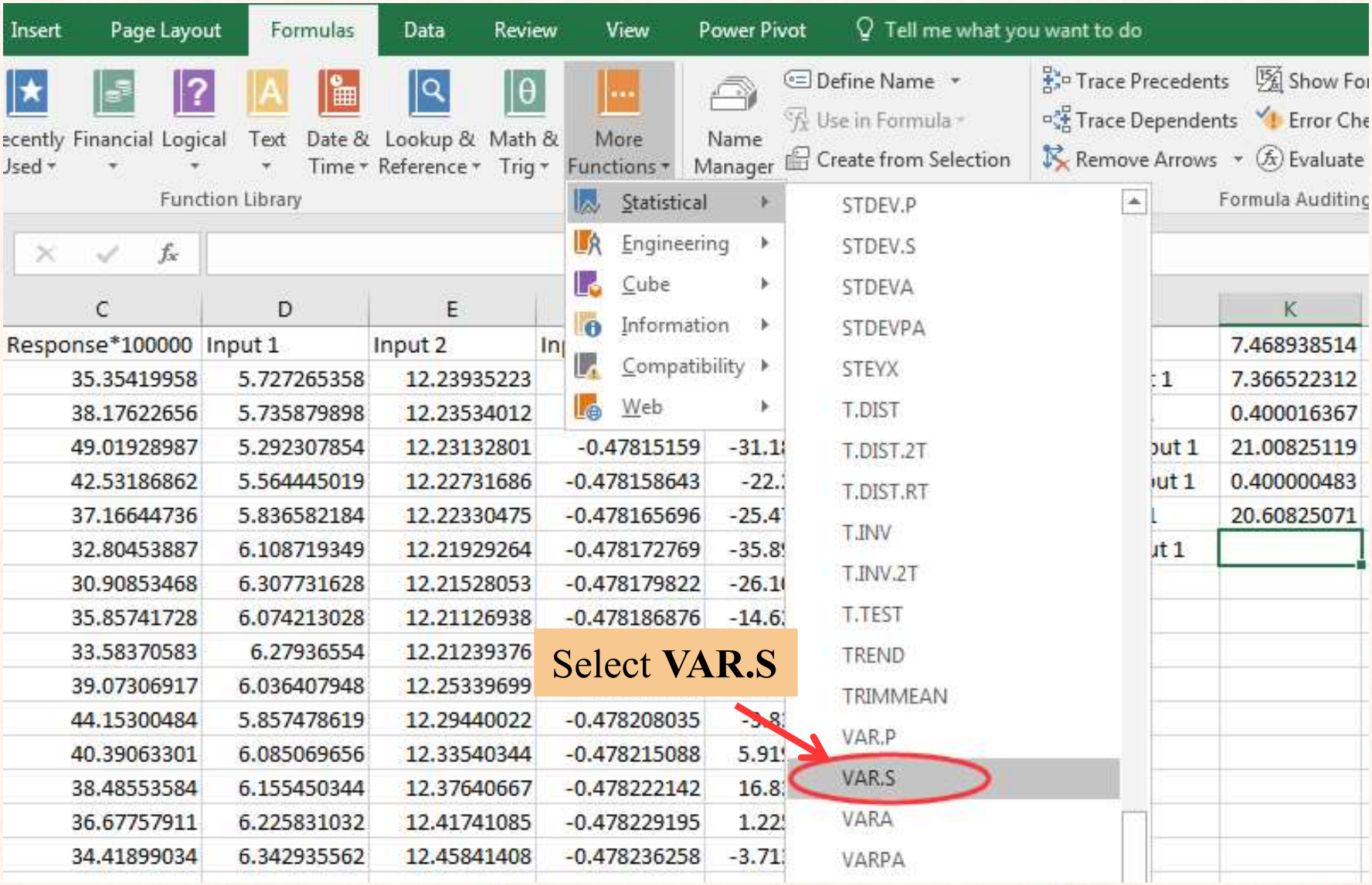
	C	D	E	F	G	H	I	J	K
	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
2	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			Mode of Input 1	0.400016367
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			Maximum of Input 1	21.00825119
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			Minimum of Input 1	0.400000483
4	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	=K4-K5
5	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
5	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
4	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
7	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
1	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
3	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
6	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				

Part A.2 Range

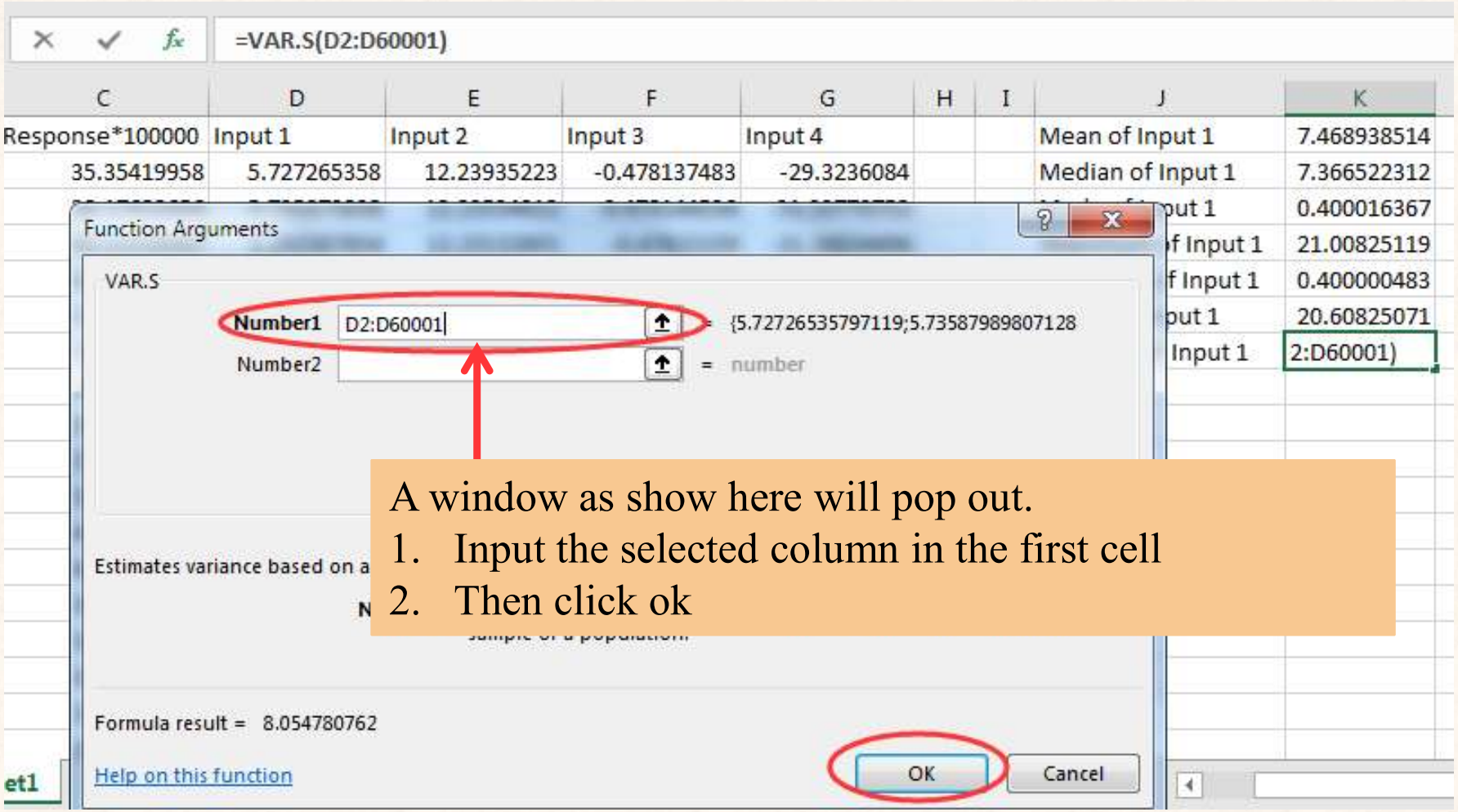


C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	20.60825071
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				

Part A.2 Sample Variance



Part A.2 Sample Variance



A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

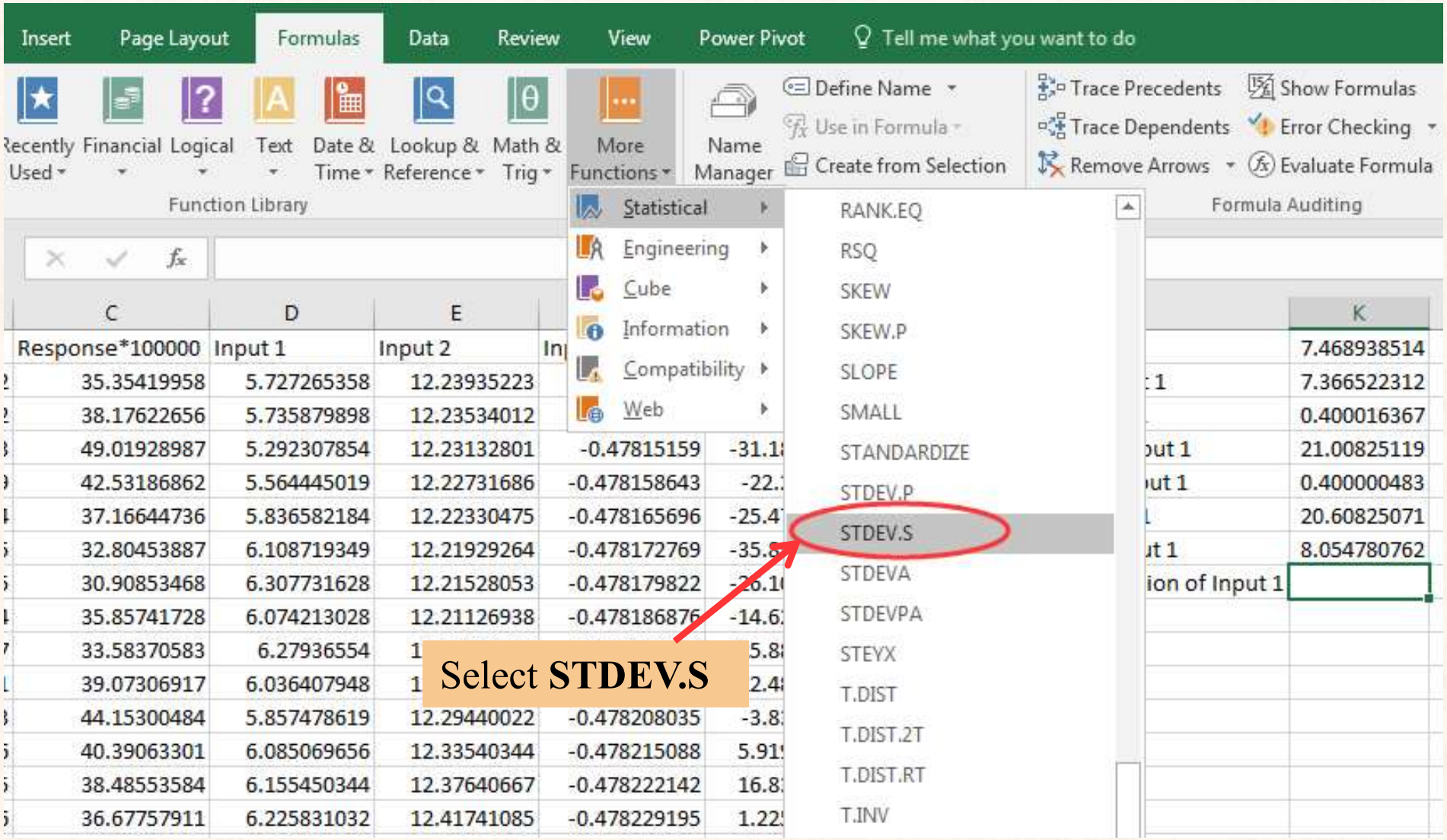
Part A.2 Sample Variance

=VAR.S(D2:D60001)

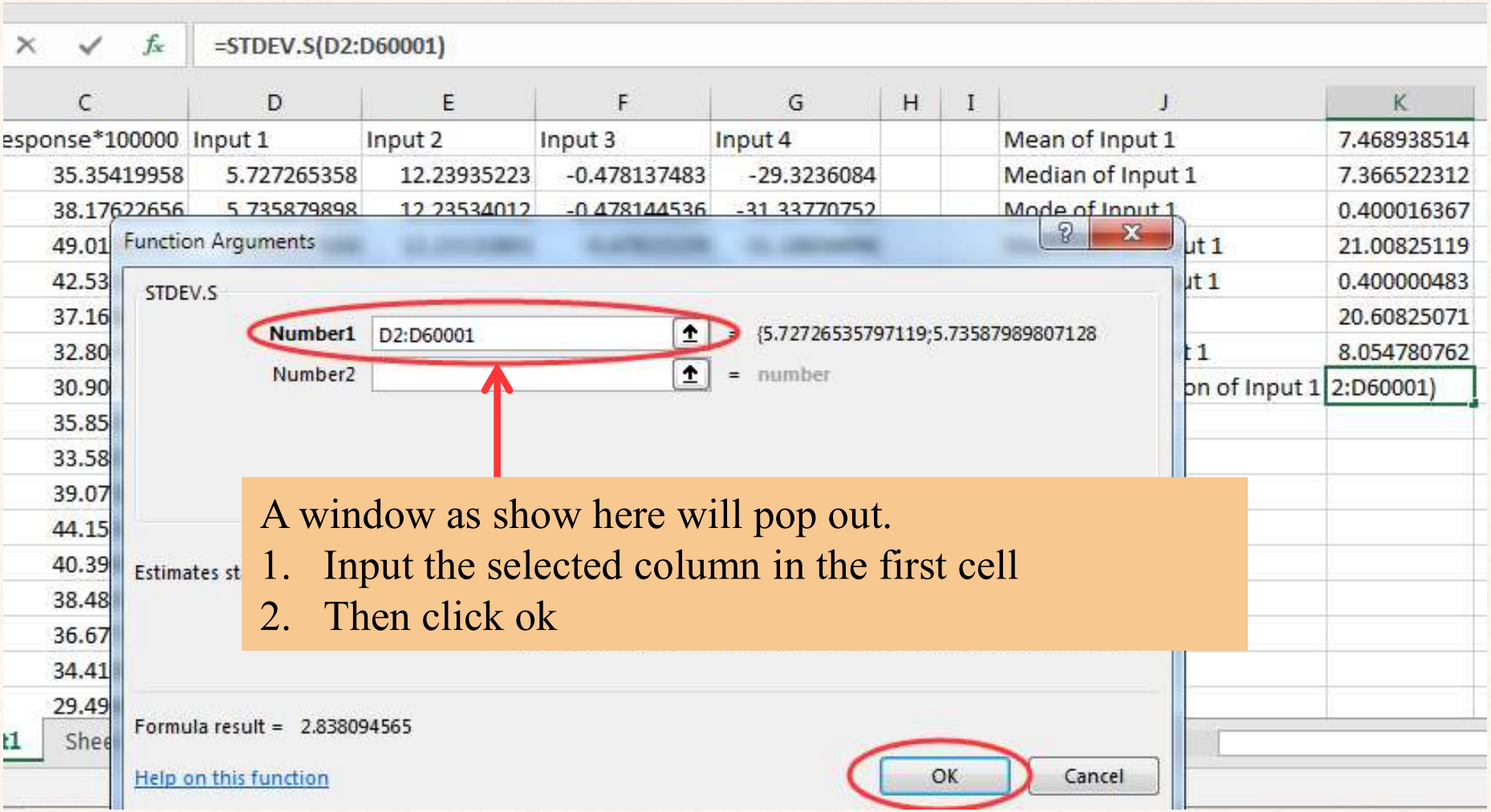
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	20.60825071
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			Variance of Input 1	8.054780762
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				

Get the final result of sample variance

Part A.2 Standard Deviation



Part A.2 Standard Deviation



Part A.2 Standard Deviation

=STDEV.S(D2:D60001)

C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	20.60825071
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			Variance of Input 1	8.054780762
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Standard deviation of Input 1	2.838094565
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				

Get the final result of standard deviation

Part B Data Representation

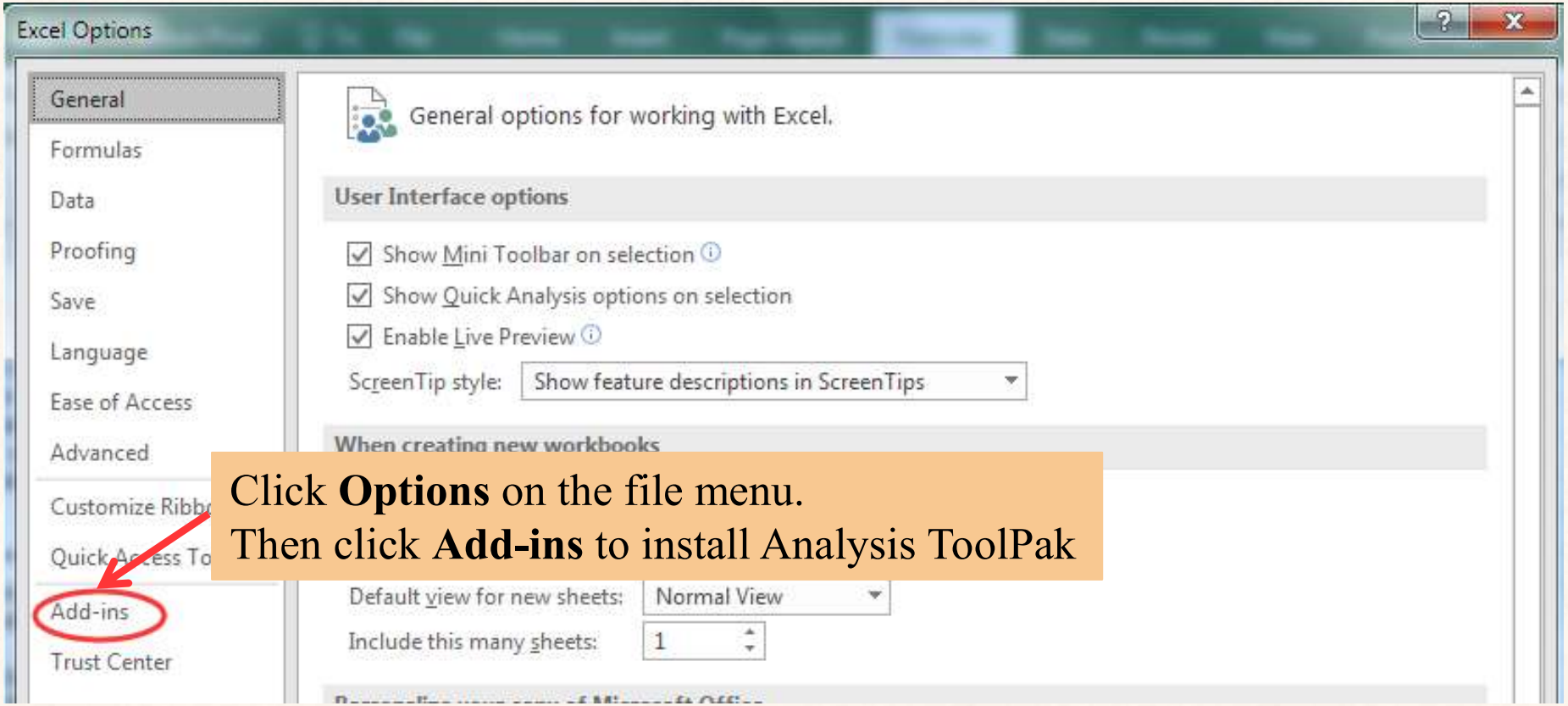
- B.1 Histograms (Frequency Histograms)

A histogram is one method of graphically representing a set of data. Graphs help to provide a sense of shape to the data points, which may provide some insight to the analysis about how the data is distributed.

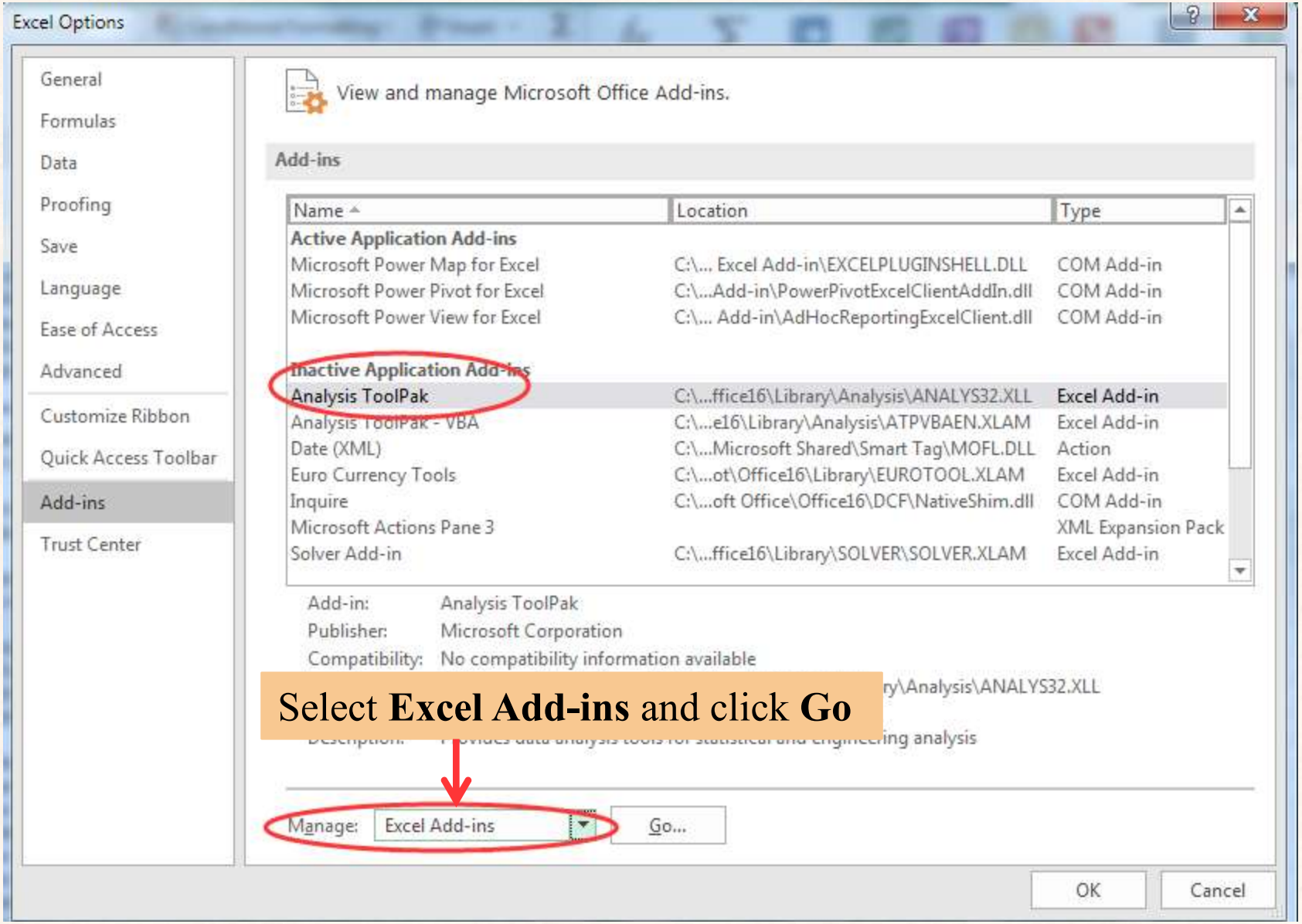
  - Create a Histogram using Data Analysis ToolPak
  - Create a Histogram using Frequency function

Part B.1 Create a Histogram using Data Analysis ToolPak

- Step 1: make sure the Analysis ToolPak Add-in is ready to use before you use the Histogram tool

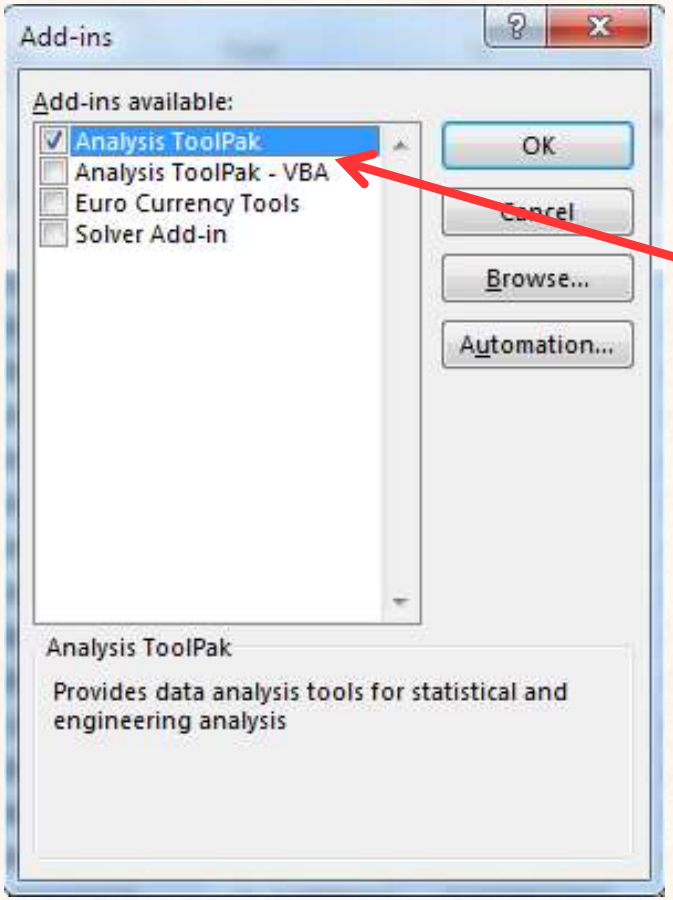


Part B.1 Create a Histogram using Data Analysis ToolPak



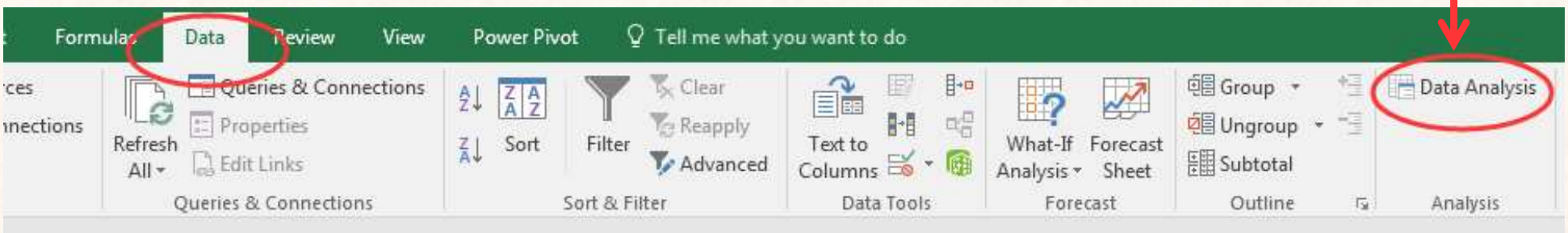
Select **Excel Add-ins** and click **Go**

Part B.1 Create a Histogram using Data Analysis ToolPak



1. In the Add-ins dialog box, select **Analysis ToolPak** and click Ok.

2. Click **Data** in the panel and check whether Analysis ToolPak has been installed



Part B.1 Create a Histogram using Data Analysis ToolPak

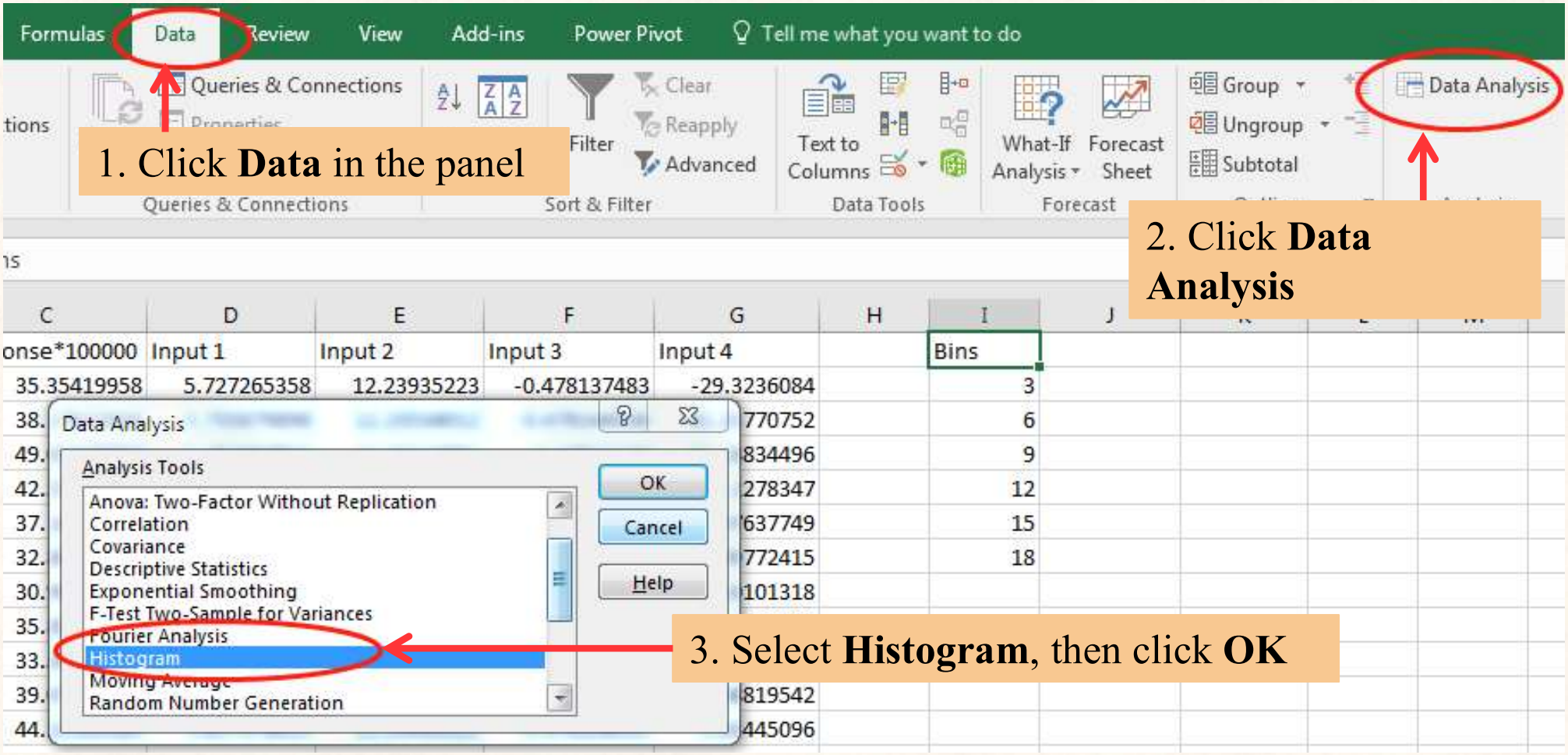
- Step 2: create a histogram using Input 1. Firstly, we need to create the data intervals in which we want to find the data frequency. There are called bins.

B	C	D	E	F	G	H	I
Response	Response*100000	Input 1	Input 2	Input 3	Input 4		Bins
0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		3
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		6
0.00049019	[illegible]	[illegible]	[illegible]	[illegible]	834496		9
0.0004253	[illegible]	[illegible]	[illegible]	[illegible]	278347		12
0.0003716	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		15
0.0003280	[illegible]	[illegible]	[illegible]	[illegible]	772415		18
0.0003090	[illegible]	[illegible]	[illegible]	[illegible]	101318		
0.0003585	[illegible]	[illegible]	[illegible]	[illegible]	547016		
0.0003358	[illegible]	[illegible]	[illegible]	[illegible]	464355		
0.000390731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		
0.00044153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096		
0.000403906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		
0.000384855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221		
0.000366776	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763		
0.00034419	34.41899034	6.342935562	12.45841408	-0.478236258	-3.713405132		
0.000294946	29.49460058	6.747737408	12.4994173	-0.478243311	-1.201598406		
0.000302216	30.22158733	6.595600128	12.54042053	-0.478250374	-0.233026966		
0.000328515	32.85147654	6.443462849	12.58142376	-0.478257428	-8.387624741		
0.000362973	36.29734264	6.260338306	12.62242699	-0.478264491	-4		
0.000412096	41.20962422	6.027213097	12.66343021	-0.478271544	-4		

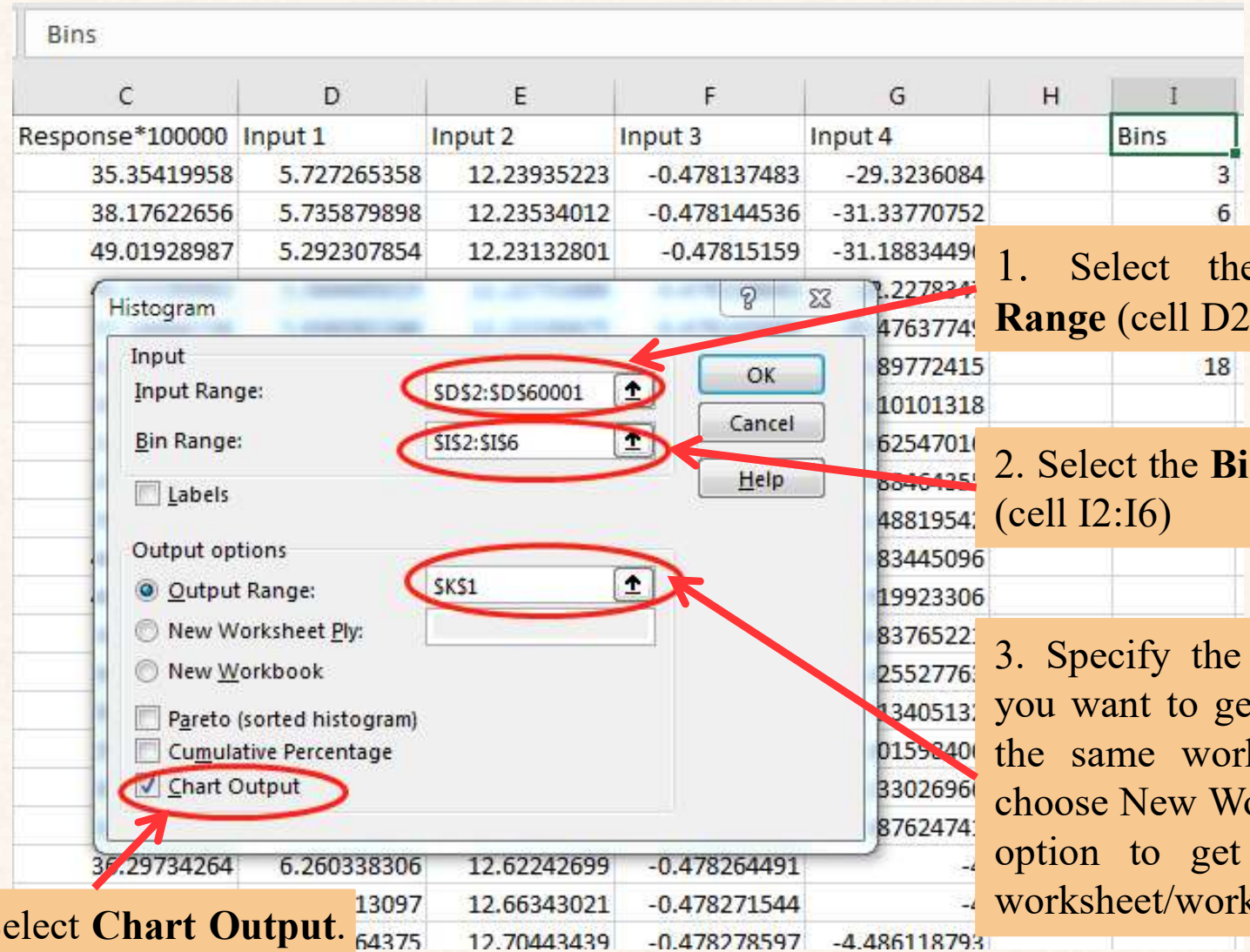
With the data Input1, you need to specify these bins separately in an additional column as shown on the right. The intervals can be determined based on the range of Input 1.

Part B.1 Create a Histogram using Data Analysis ToolPak

- Step 3: select Histogram in the Data dialog box to create a histogram based on the data, Input 1 and created bins.



Part B.1 Create a Histogram using Data Analysis ToolPak



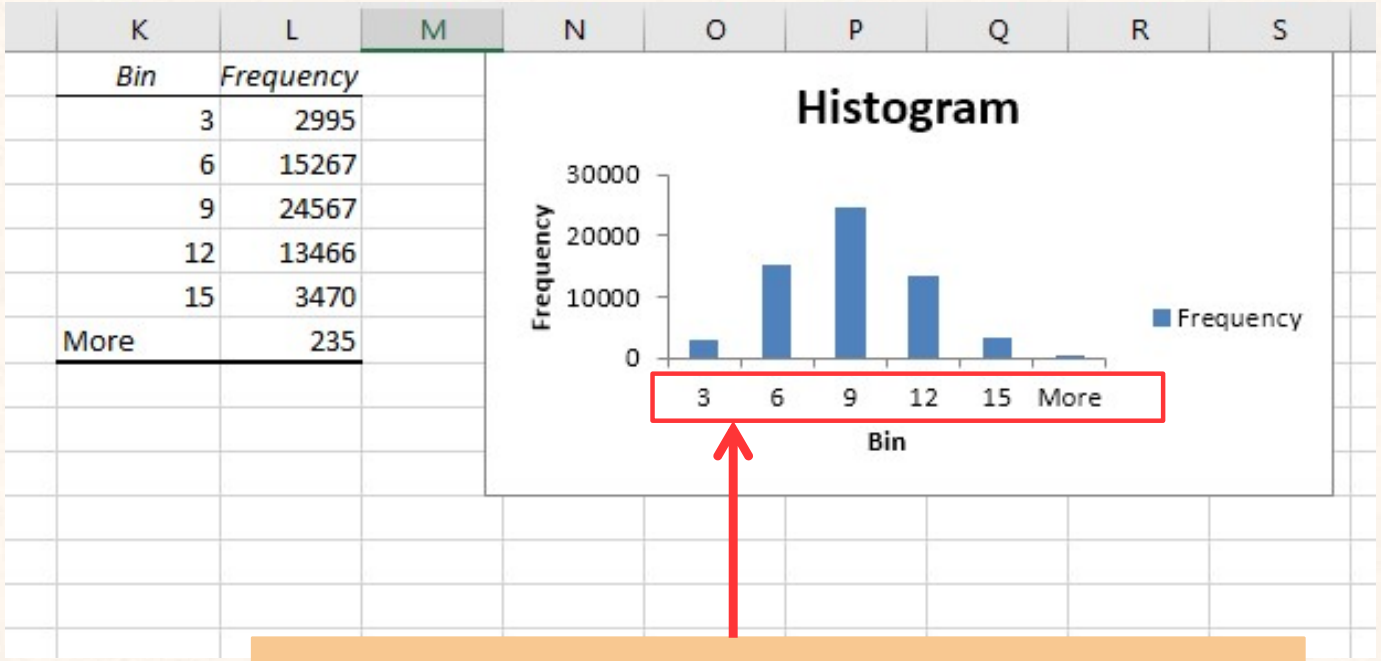
1. Select the **Input Range** (cell D2:D60001)

2. Select the **Bin Range** (cell I2:I6)

3. Specify the **Output Range** if you want to get the Histogram in the same worksheet. Otherwise, choose New Worksheet/Workbook option to get it in a separate worksheet/workbook

4. Select **Chart Output**. Then click OK

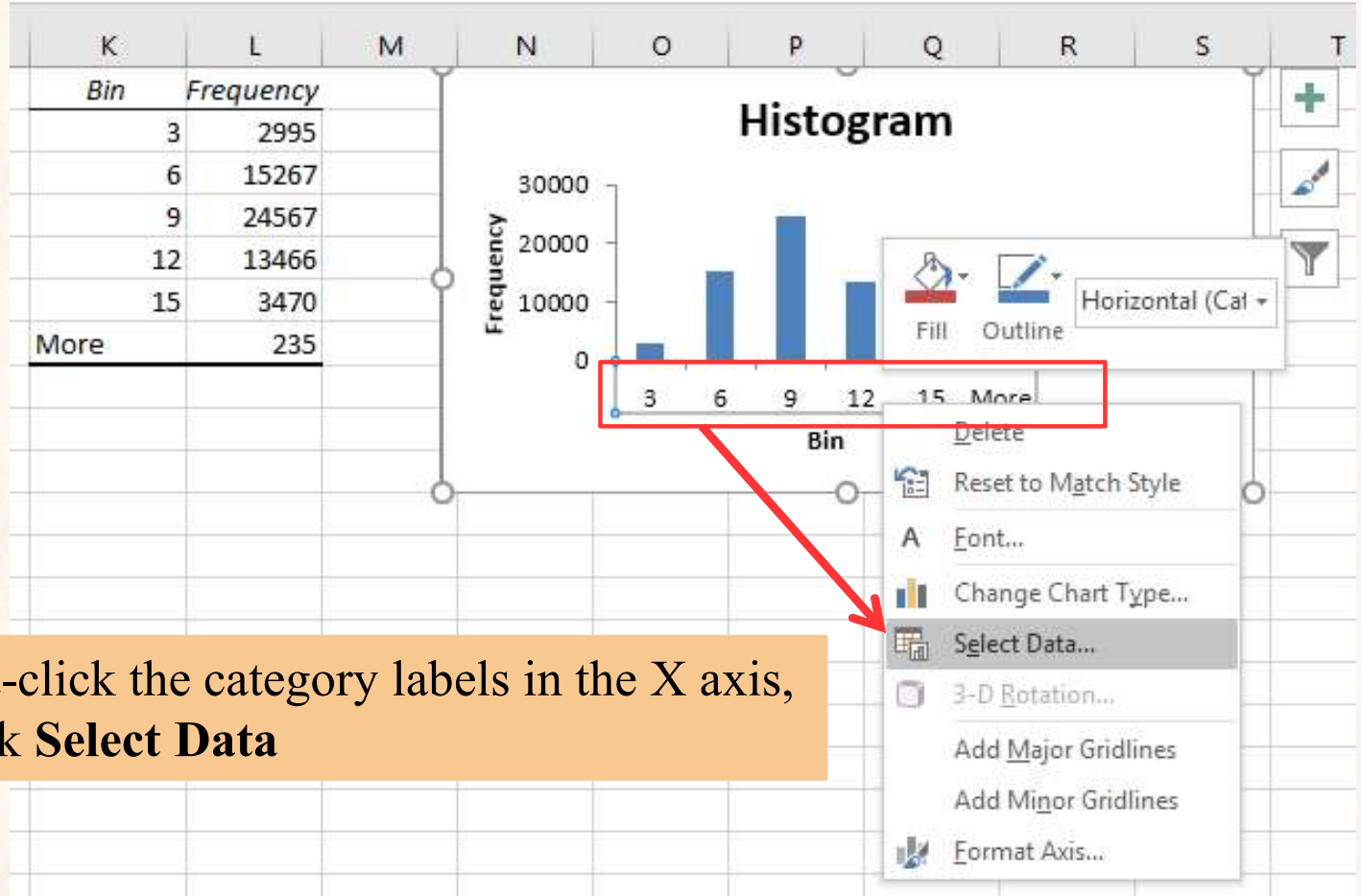
Part B.1 Create a Histogram using Data Analysis ToolPak



With the Analysis ToolPak, Exel adds the horizontal axis labels based on the bin numbers that you specify. How to display ranges instead of bin numbers?

Part B.1 Create a Histogram using Data Analysis ToolPak

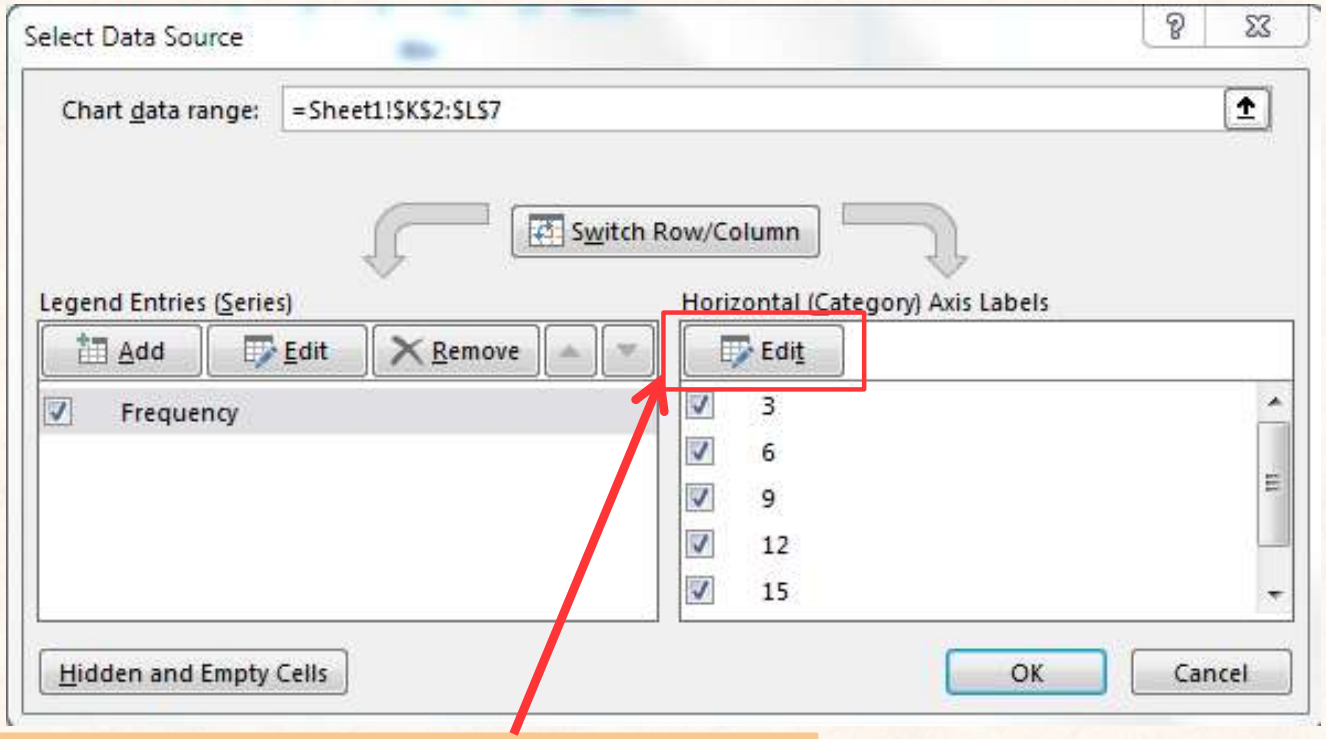
- Change the axis labels on an Excel histogram chart



1. Right-click the category labels in the X axis, and click **Select Data**

Part B.1 Create a Histogram using Data Analysis ToolPak

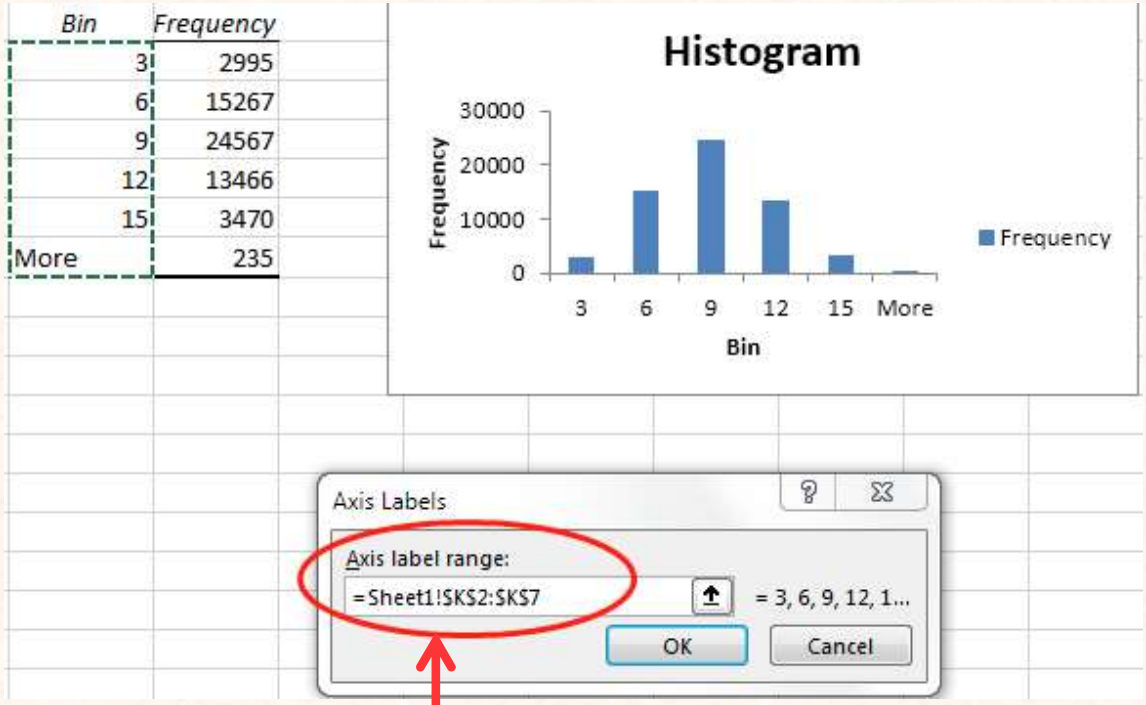
- Change the axis labels on an Excel histogram chart

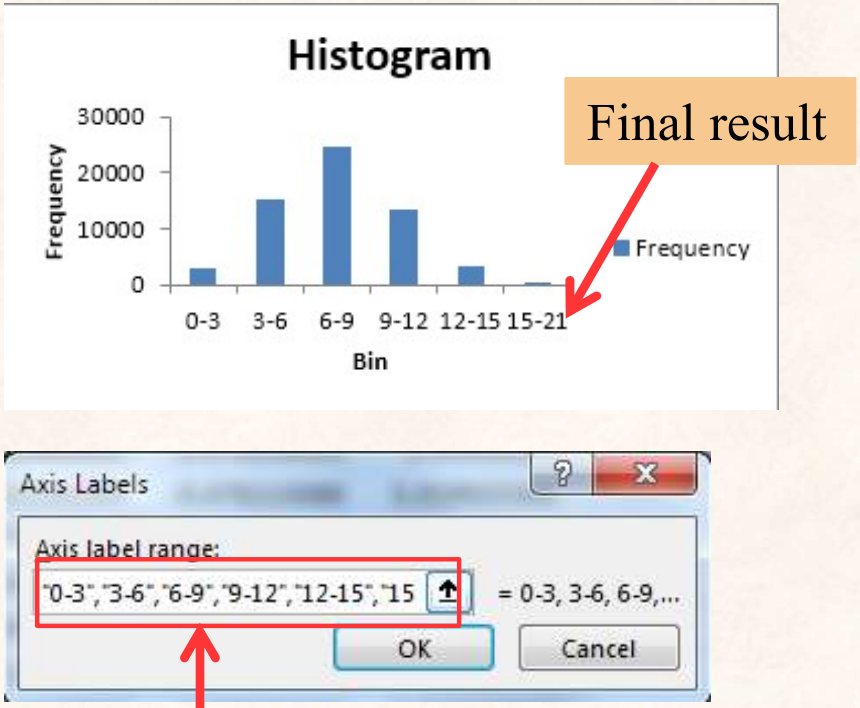


2. On the right-hand side pane, under Horizontal (category) Axis Labels, click the **Edit** button

Part B.1 Create a Histogram using Data Analysis ToolPak

- Change the axis labels on an Excel histogram chart





3. Enter the labels you want to display, separated by commas. If you are entering the intervals, enclose them in double quotes.

4. After entering the intervals, click **OK**

Part B.1 Create a Histogram using formulas

- FREQUENCY function:

It's the other way to create a histogram in Excel. The biggest advantage of using FREQUENCY function is that you won't have to re-do your histogram with each change in the input data. The histogram will **update automatically** as soon as you edit, add new or delete existing input values.

- Step 1: select your source data and create the bin numbers in a new column.

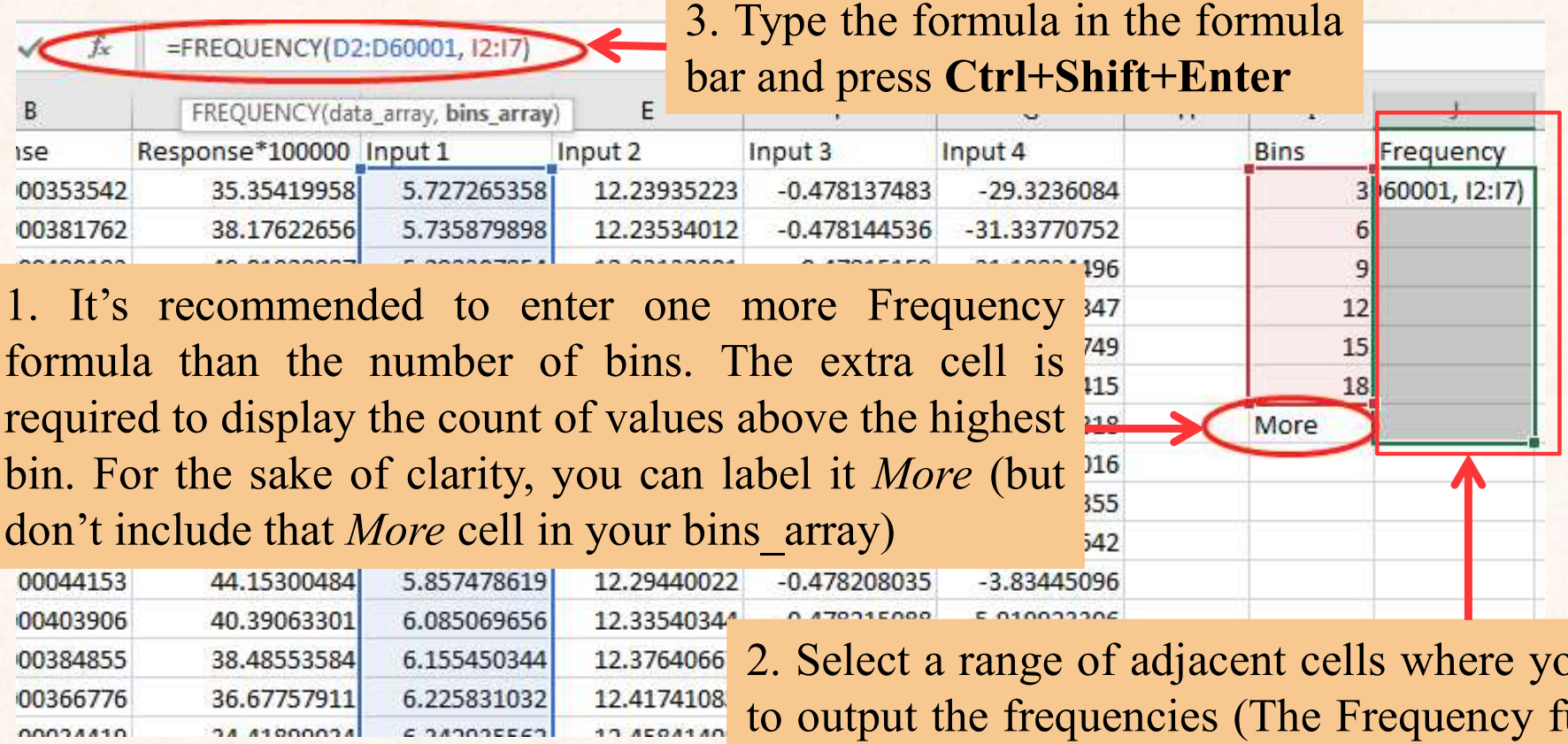
B	C	D	E	F	G	H	I
Response	Response*100000	Input 1	Input 2	Input 3	Input 4		Bins
0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		3
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		6
[illegible]	[illegible]	5.292307854	12.[illegible]	[illegible]	[illegible]		9
[illegible]	[illegible]	5.564445019	12.[illegible]	[illegible]	[illegible]		12
0.000371664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		15
0.000328045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415		18
0.000309085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318		
0.000358574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		

Select source data Input 1

Create the bin numbers

Part B.1 Create a Histogram using formulas

- Step 2: using function **FREQUENCY(data_array, bins_array)**
  data_array – a set of values for which you want to count frequencies.
  bins_array – a set of bins for grouping the values.



3. Type the formula in the formula bar and press **Ctrl+Shift+Enter**

1. It's recommended to enter one more Frequency formula than the number of bins. The extra cell is required to display the count of values above the highest bin. For the sake of clarity, you can label it *More* (but don't include that *More* cell in your bins_array)

2. Select a range of adjacent cells where you want to output the frequencies (The Frequency function should be entered as a **multi-cell array formula**).

Part B.1 Create a Histogram using formulas

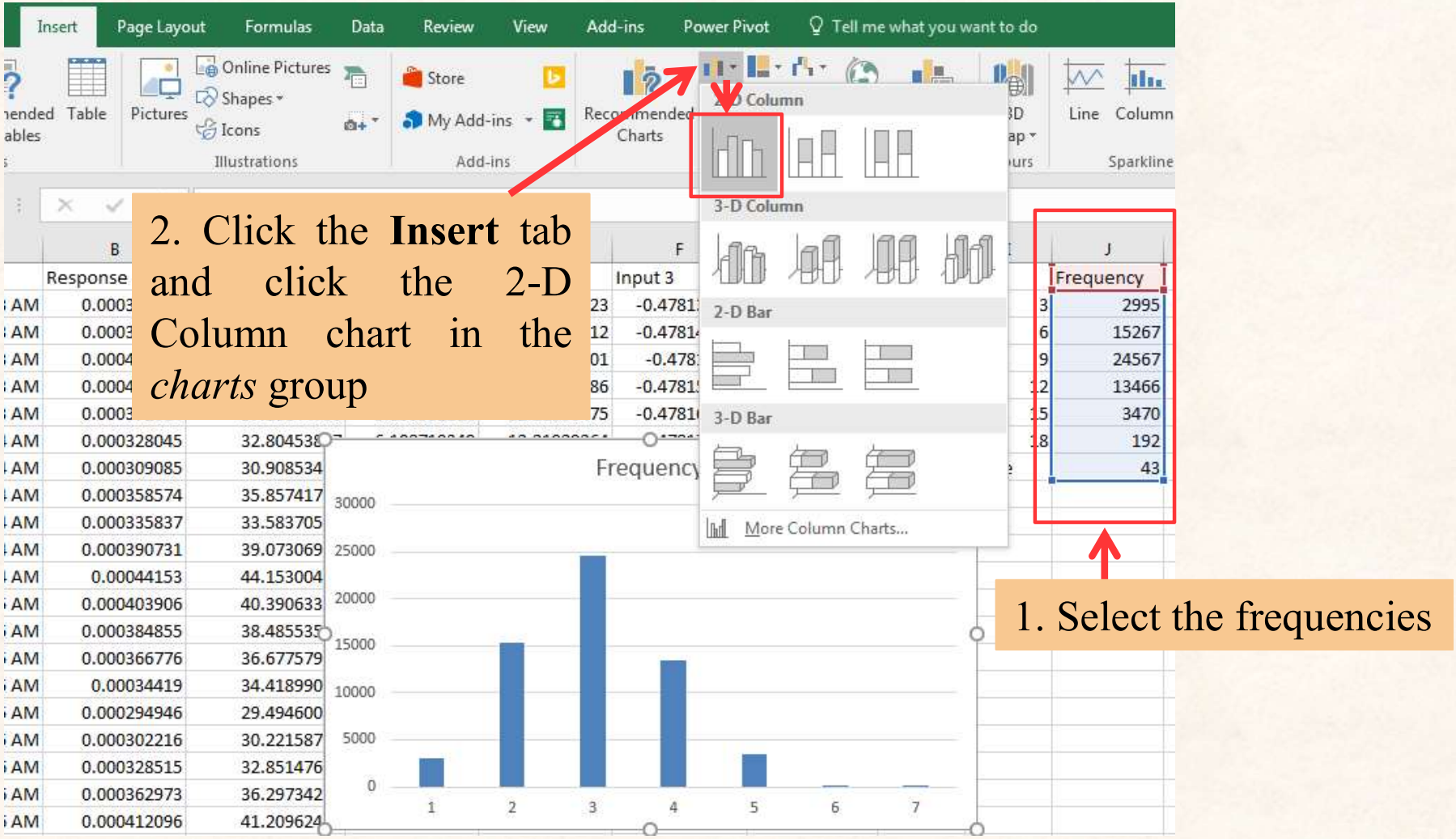
Get the results of Frequency. The Excel FREQUENCY function returns values that are greater than a previous bin and less than or equal to a given bin. The last Frequency formula returns the number of values greater than the highest bin.

	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4		Bins	Frequency
2	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		3	2995
2	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		6	15267
3	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		9	24567
9	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		12	13466
4	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		15	3470
5	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415		18	192
5	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318		More	43
4	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016			
7	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355			
1	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542			
3	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096			
6	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306			
5	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			
6	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763			

H	I	J
Interval	Bins	Frequency
0-3	3	2995
3-6	6	15267
6-9	9	24567
9-12	12	13466
12-15	15	3470
15-18	18	192
Over 18	More	43

Part B.1 Create a Histogram using formulas

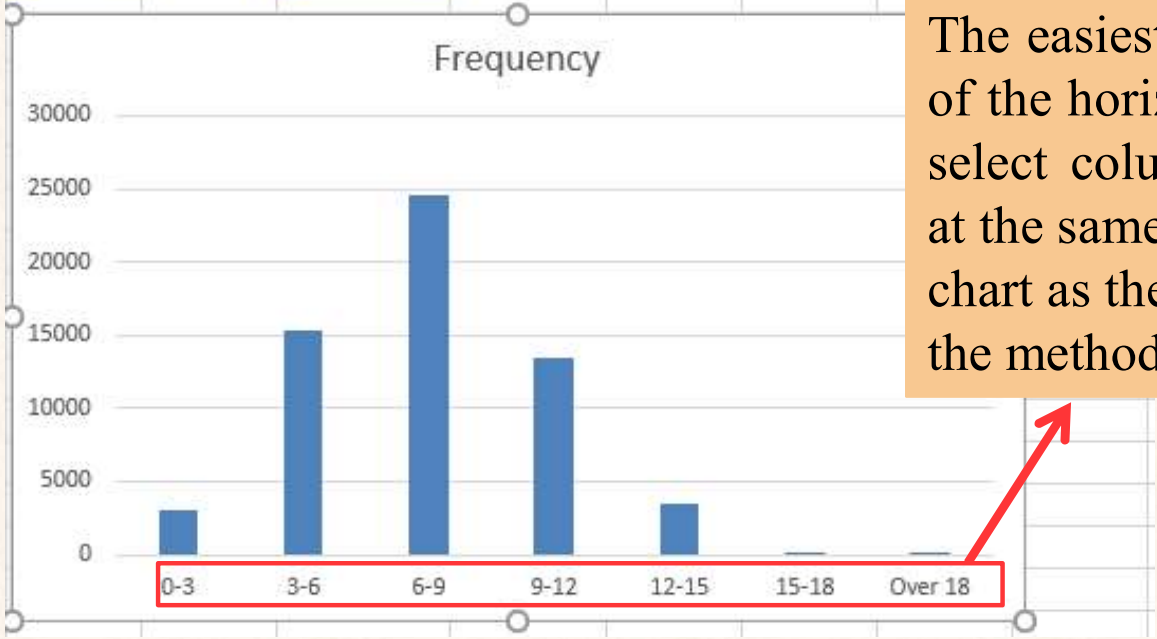
- Step 3: make a histogram based on the summary data



Part B.1 Create a Histogram using formulas

- Step 3: make a histogram based on the summary data

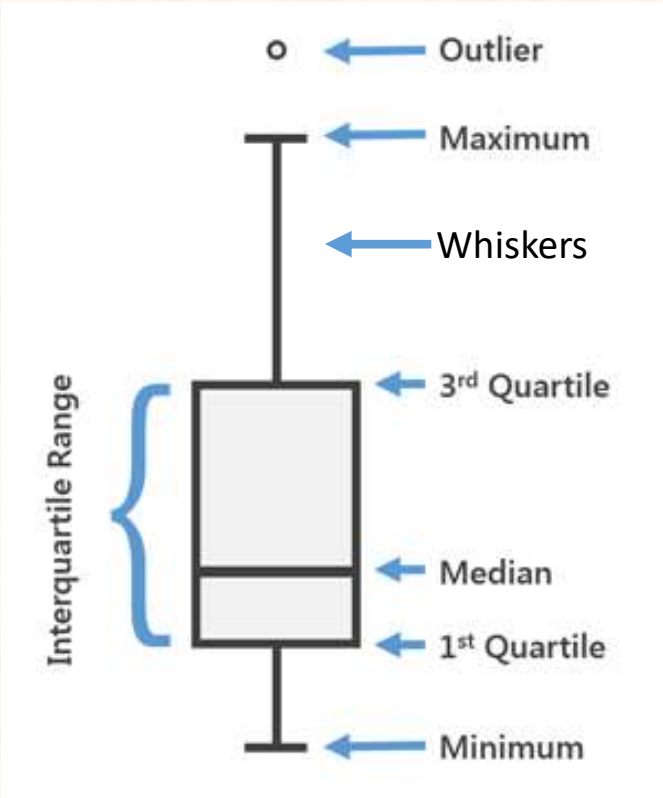
H	I	J
Interval	Bins	Frequency
0-3	3	2995
3-6	6	15267
6-9	9	24567
9-12	12	13466
12-15	15	3470
15-18	18	192
Over 18	More	43



The easiest way to change the default labels of the horizontal axis with bin numbers is to select columns of **Interval** and **Frequency** at the same time. Then click the 2-D Column chart as the previous slide. Also, you can use the method in P40.

Part B.2 Box Plot (Box-whisker Plot)

- In a box plot, numerical data is divided into quartiles, and a box is drawn between the first and third quartiles, with an additional line drawn along the second quartile to mark the median. The minimum and maximum outside the first and third quartiles are depicted with lines/whiskers.



Part B.2 Box Plot (Box-whisker Plot)

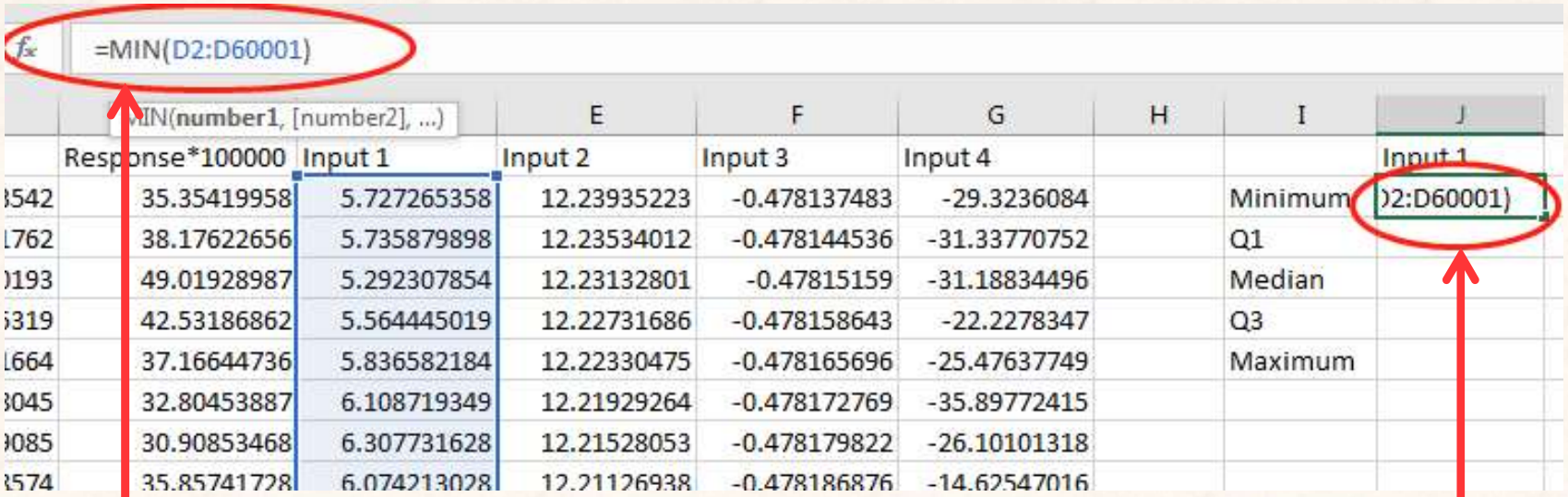
- Step 1: Calculate the quartile values

Value	Formula
Minimum value	MIN(cell range)
First quartile	QUARTILE.INC(cell range, 1)
Median value	QUARTILE.INC(cell range, 2)
Third quartile	QUARTILE.INC(cell range, 3)
Maximum value	MAX(cell range)

How to calculate the five values? You can have try it by yourself with the method we have learned in Part A.

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values
  Since the formula is given, the easiest way to calculate the value is to use the formula in the formula bar directly.



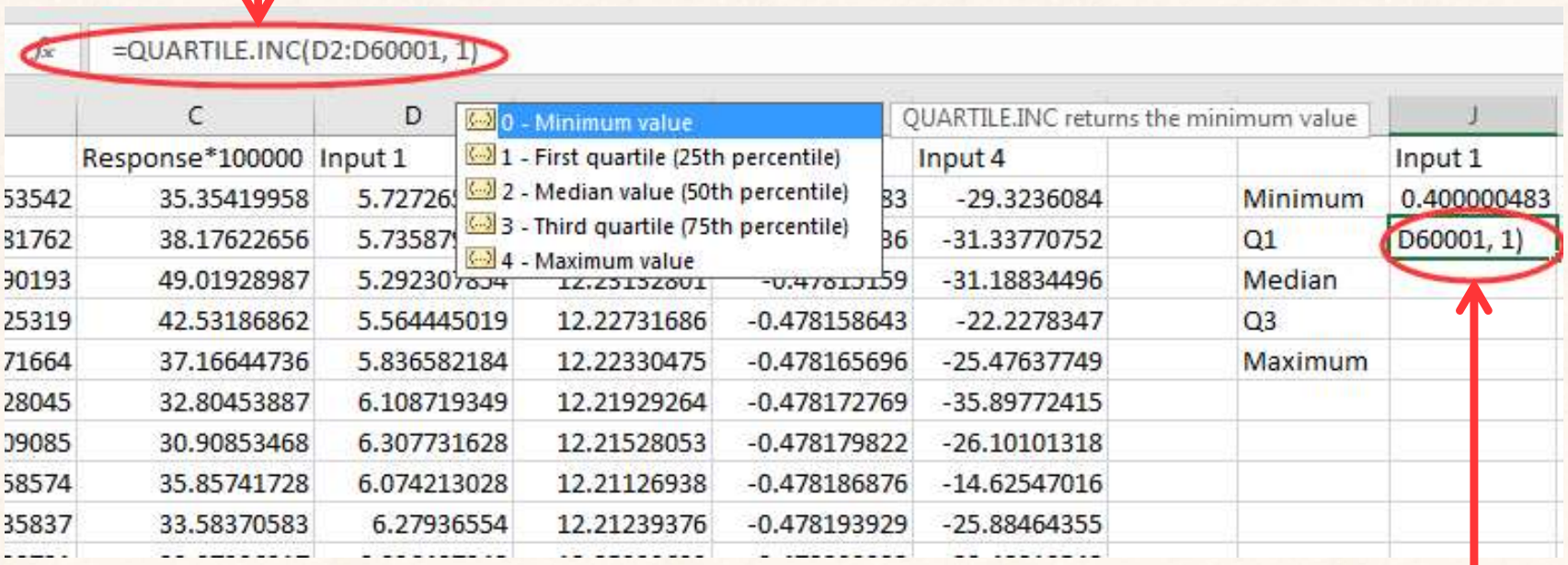
1. place your cursor in the cell labeled Minimum

2. Type "=MIN(cell range)" and press Enter

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type "=QUARTILE.INC(cell range, 1)" and press Enter

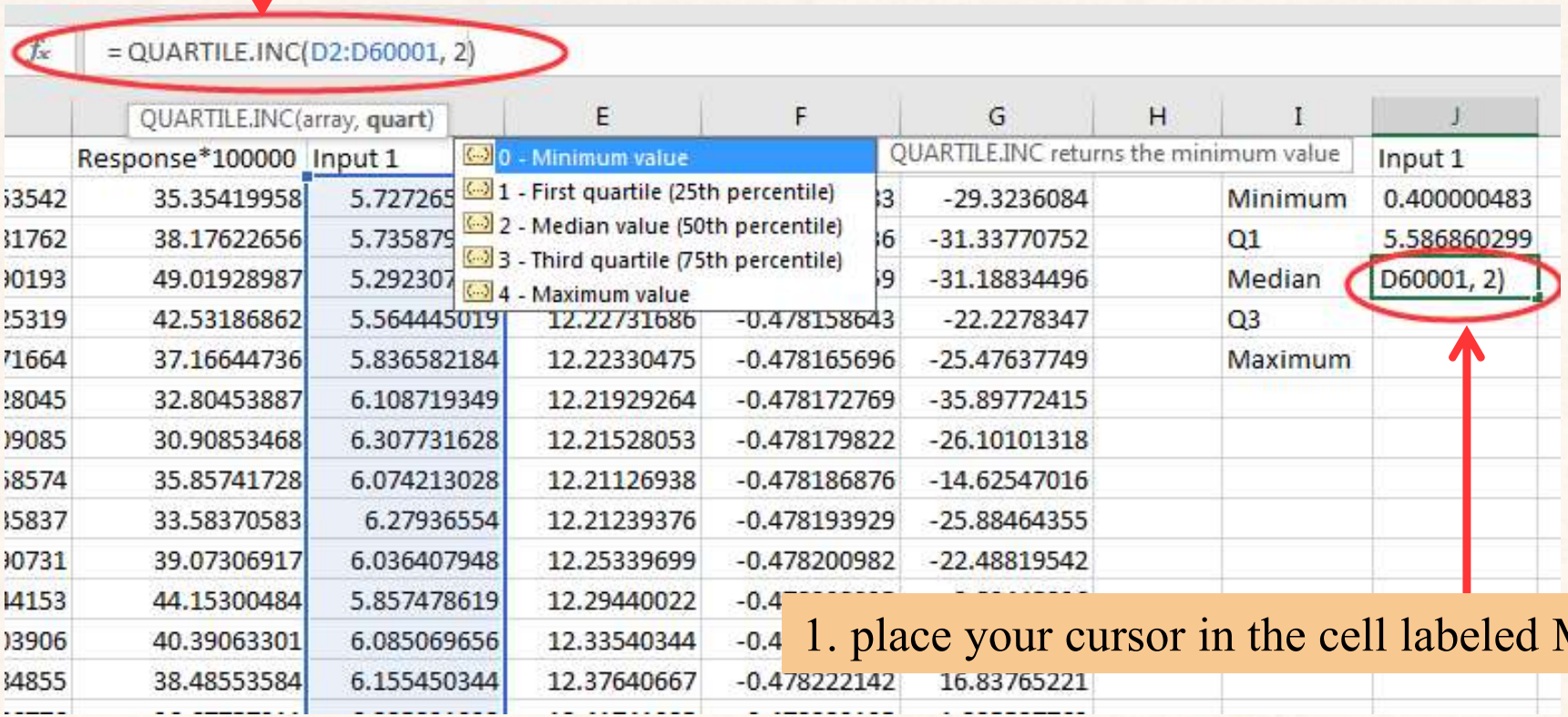


1. place your cursor in the cell labeled Q1

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type "=QUARTILE.INC(cell range, 2)" and press Enter

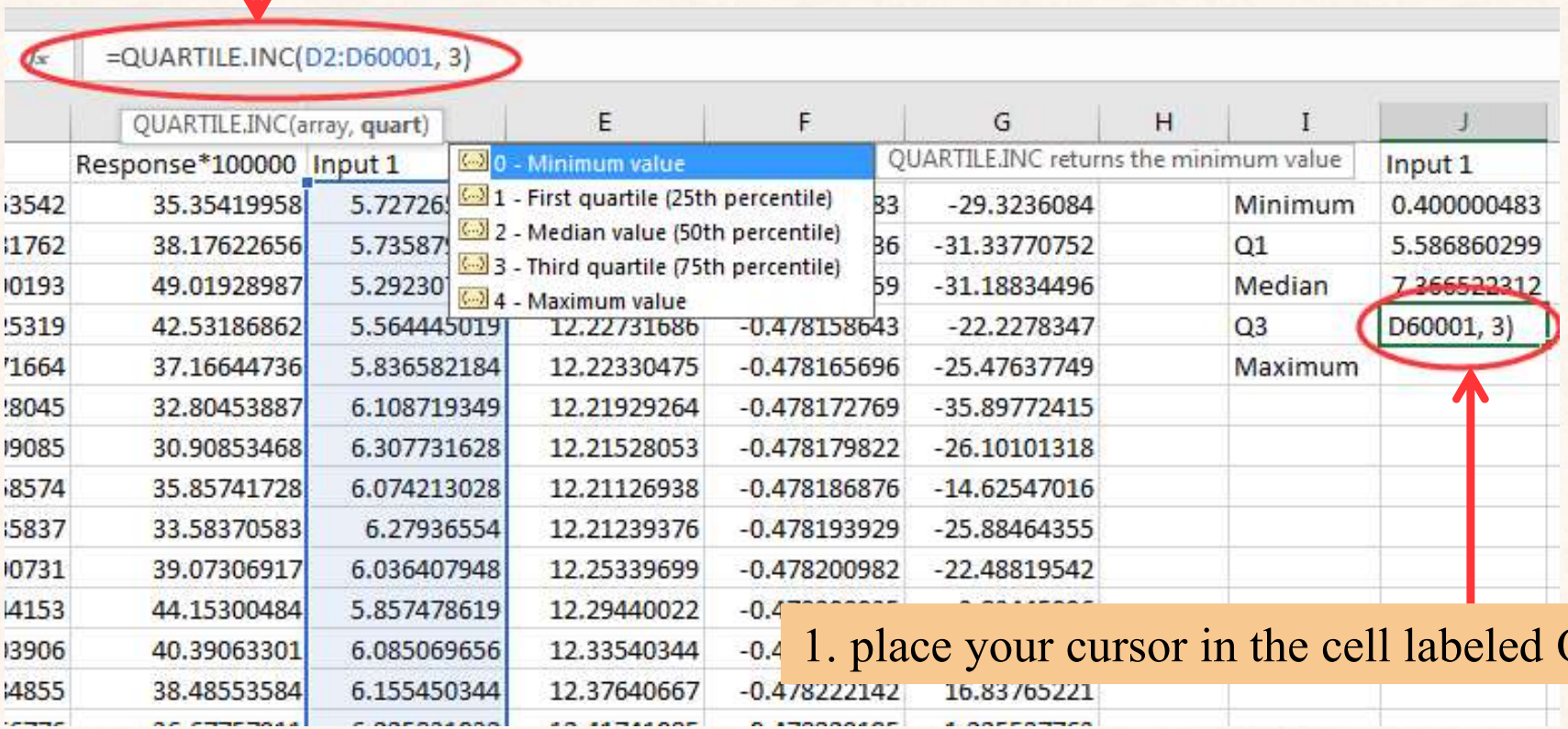


1. place your cursor in the cell labeled Median

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type "=QUARTILE.INC(cell range, 3)" and press Enter

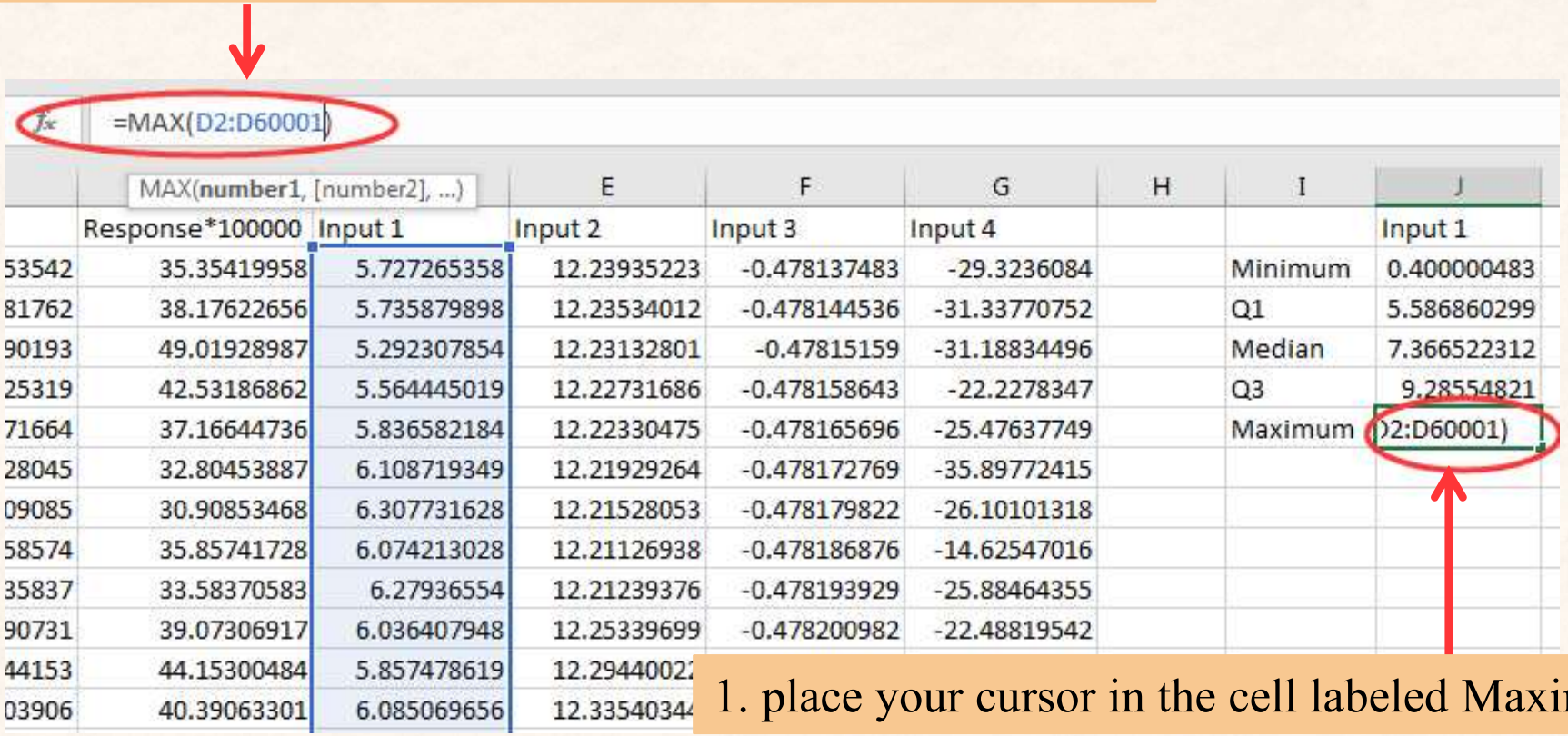


1. place your cursor in the cell labeled Q3

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type "=Maximum(cell range)" and press Enter



	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
53542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		Minimum	0.400000483
81762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		Q1	5.586860299
90193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		Median	7.366522312
25319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
71664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	
28045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
09085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			
58574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016			
35837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355			
90731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542			
44153	44.15300484	5.857478619	12.2944002					
03906	40.39063301	6.085069656	12.3354034					

1. place your cursor in the cell labeled Maximum

Part B.2 Box Plot (Box-whisker Plot)

- Step 2: Calculate quartile differences as the following
  - First quartile – Minimum
  - Median – first quartile
  - Third quartile – mean
  - Maximum – third quartile

fx =J2

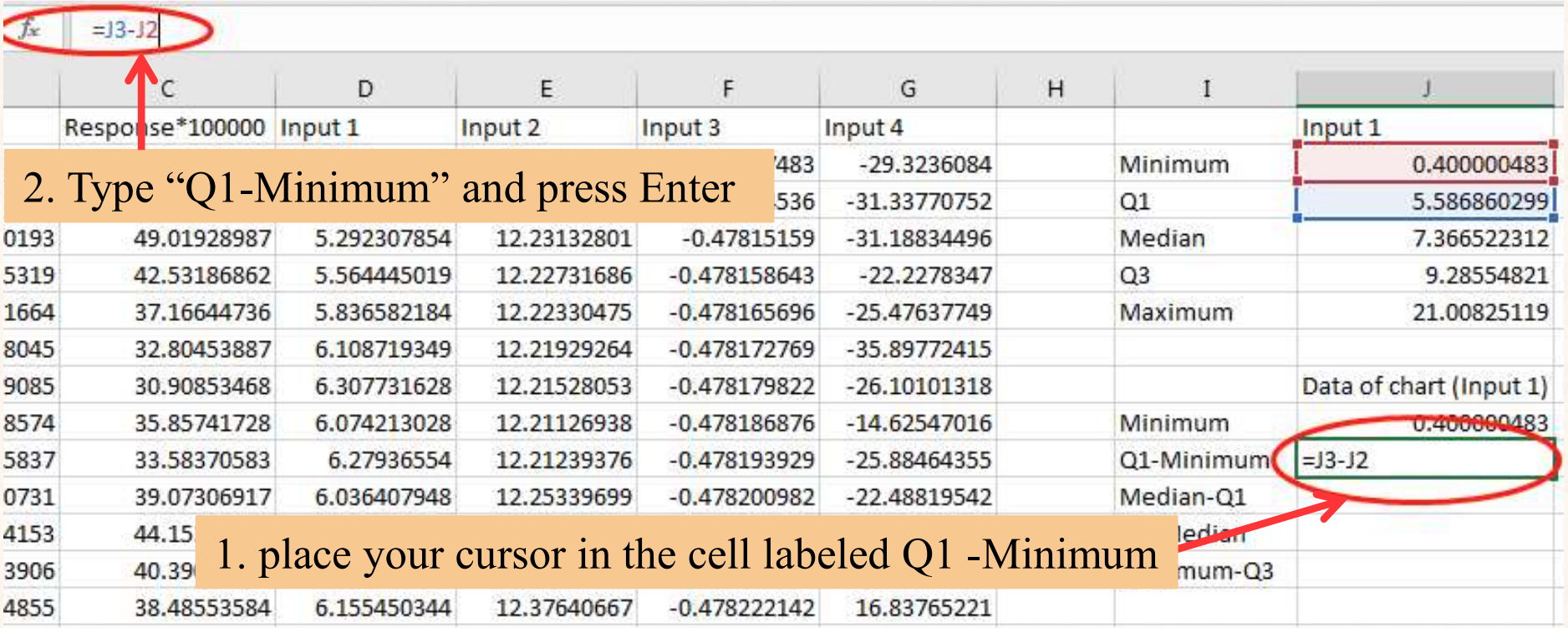
	C	D	E	F	G	H	I	J
	Respon*100000	Input 1	Input 2	Input 3	Input 4			Input 1
3542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		Minimum	0.400000483
1762	[illegible]	[illegible]	[illegible]	[illegible]536	-31.33770752		Q1	5.586860299
01	[illegible]	[illegible]	[illegible]	[illegible]59	-31.18834496		Median	7.366522312
5319	42.53186862	5.504445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
1664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
8045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
9085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	=J2
5837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	
07							Median-Q1	
41							Q3-Median	
3906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		Maximum-Q3	

2. Type “Minimum” and press Enter

1. To create the dataset of the chat, paste Minimum firstly.

Part B.2 Box Plot (Box-whisker Plot)

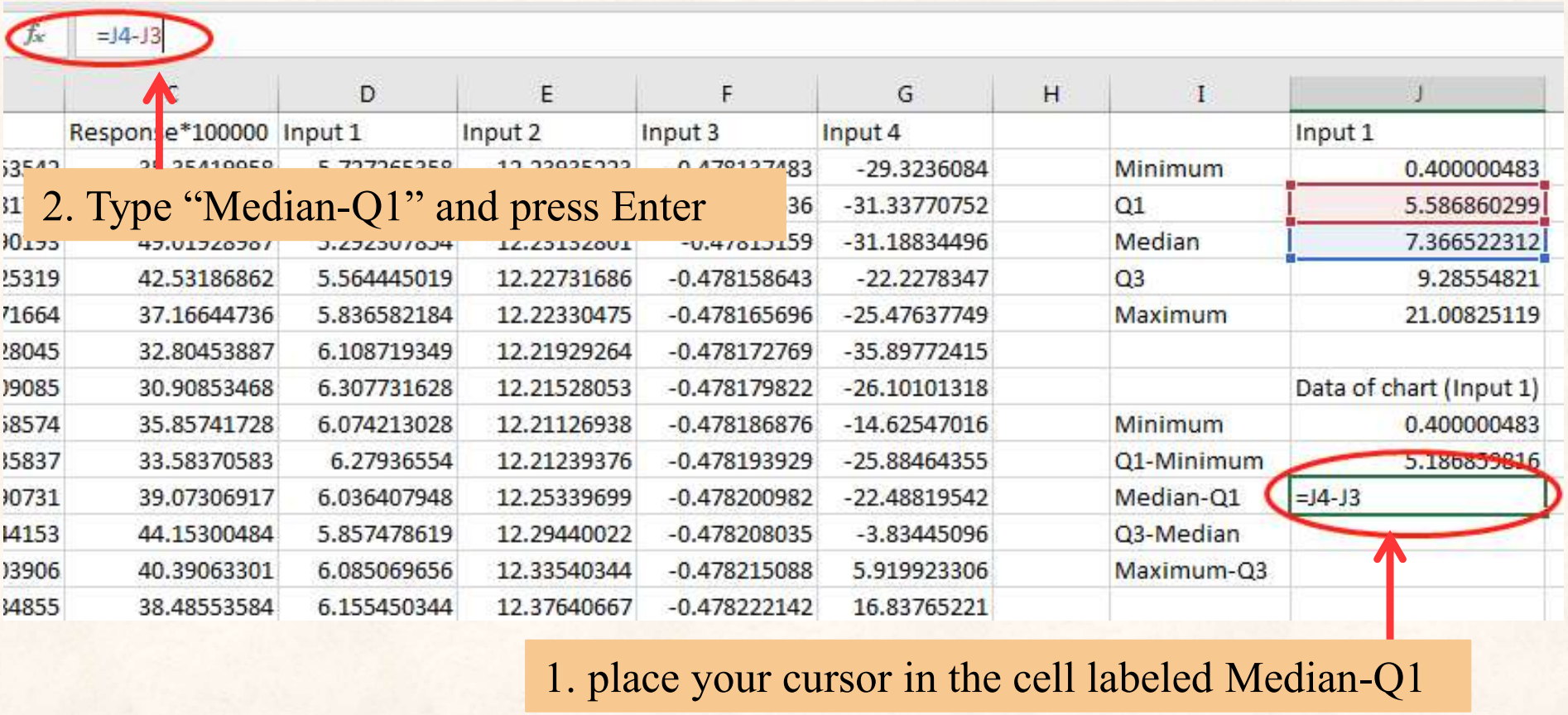
- Step 2: Calculate quartile differences as the following



	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
				483	-29.3236084		Minimum	0.400000483
				536	-31.33770752		Q1	5.586860299
0193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		Median	7.366522312
5319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
1664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
8045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
9085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	0.400000483
5837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	=J3-J2
0731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		Median-Q1	
4153	44.15							
3906	40.39						mum-Q3	
4855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			

Part B.2 Box Plot (Box-whisker Plot)

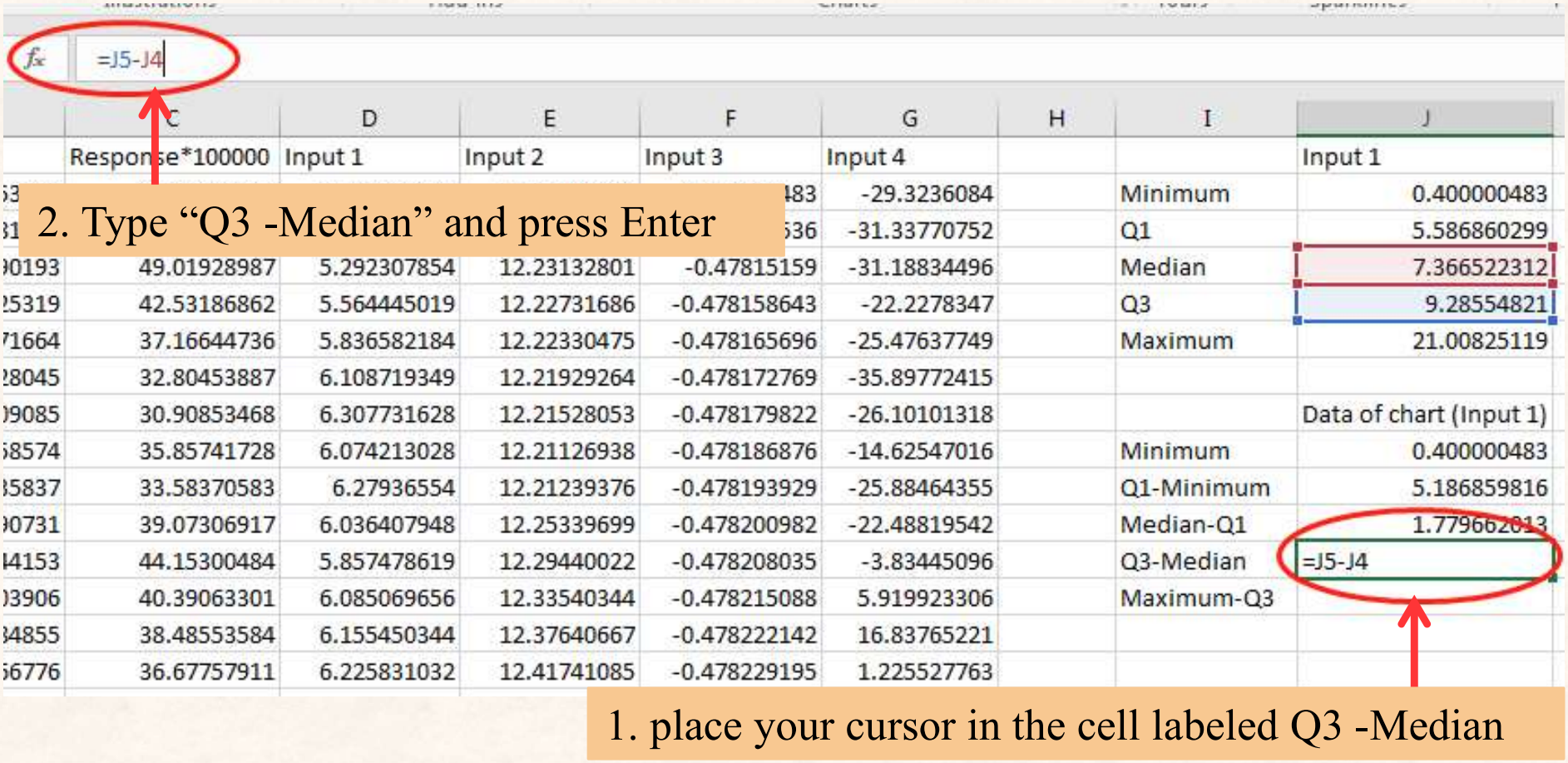
- Step 2: Calculate quartile differences



	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
					-29.3236084		Minimum	0.400000483
					-31.33770752		Q1	5.586860299
					-31.18834496		Median	7.366522312
25319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
71664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
28045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
09085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
58574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	0.400000483
35837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	5.186859816
90731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		Median-Q1	=J4-J3
44153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096		Q3-Median	
03906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		Maximum-Q3	
84855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			

Part B.2 Box Plot (Box-whisker Plot)

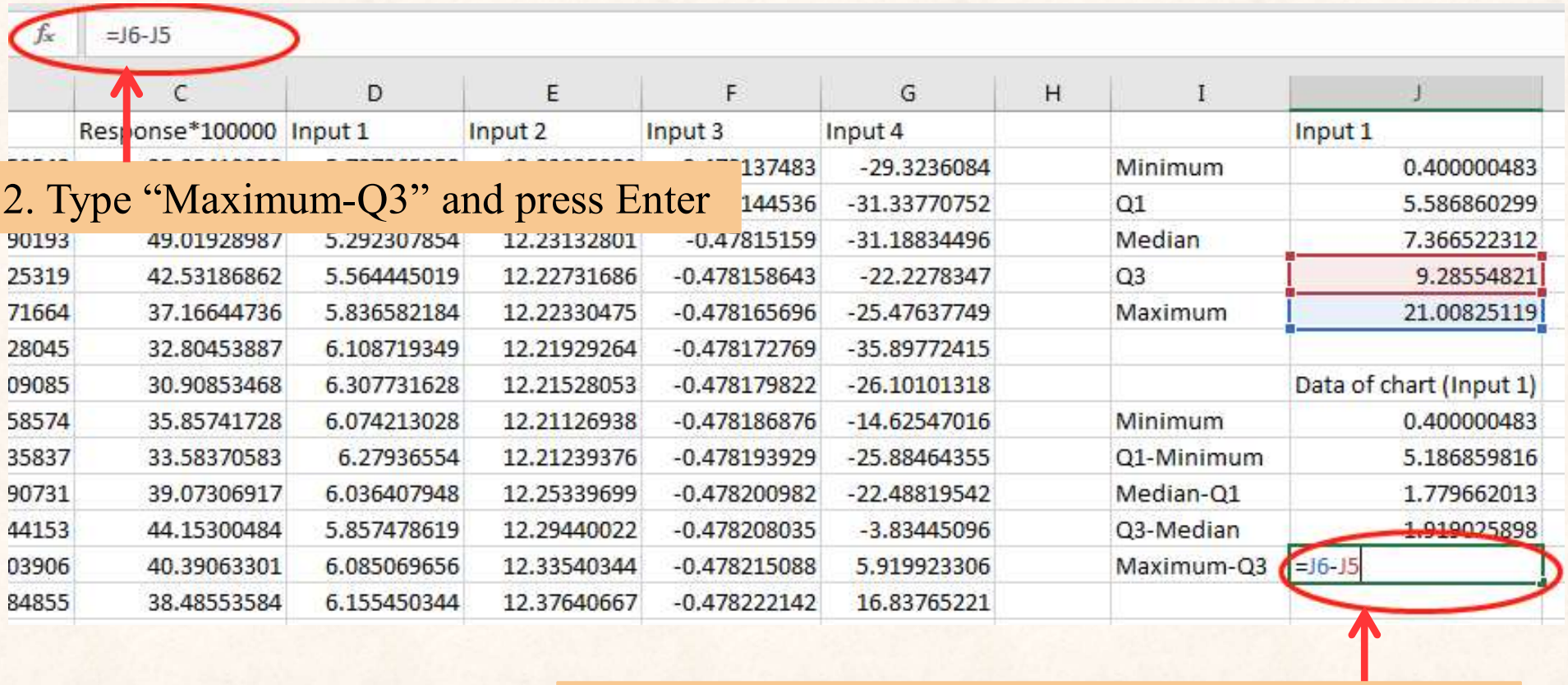
- Step 2: Calculate quartile differences



	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
53				483	-29.3236084		Minimum	0.400000483
31				536	-31.33770752		Q1	5.586860299
0193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		Median	7.366522312
25319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
71664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
28045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
09085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
58574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	0.400000483
35837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	5.186859816
0731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		Median-Q1	1.779662013
4153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096		Q3-Median	=J5-J4
03906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		Maximum-Q3	
34855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			
56776	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763			

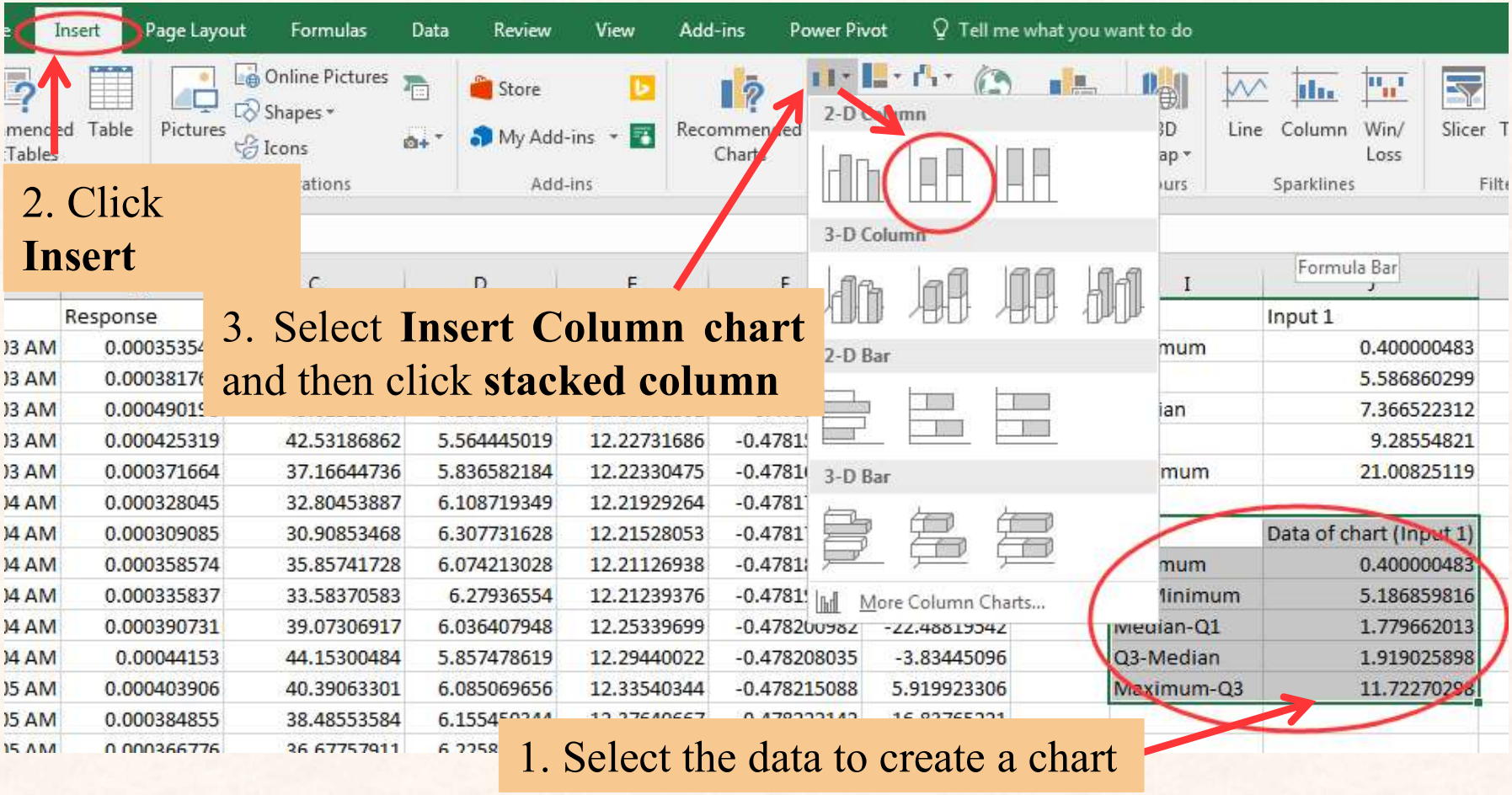
Part B.2 Box Plot (Box-whisker Plot)

- Step 2: Calculate quartile differences



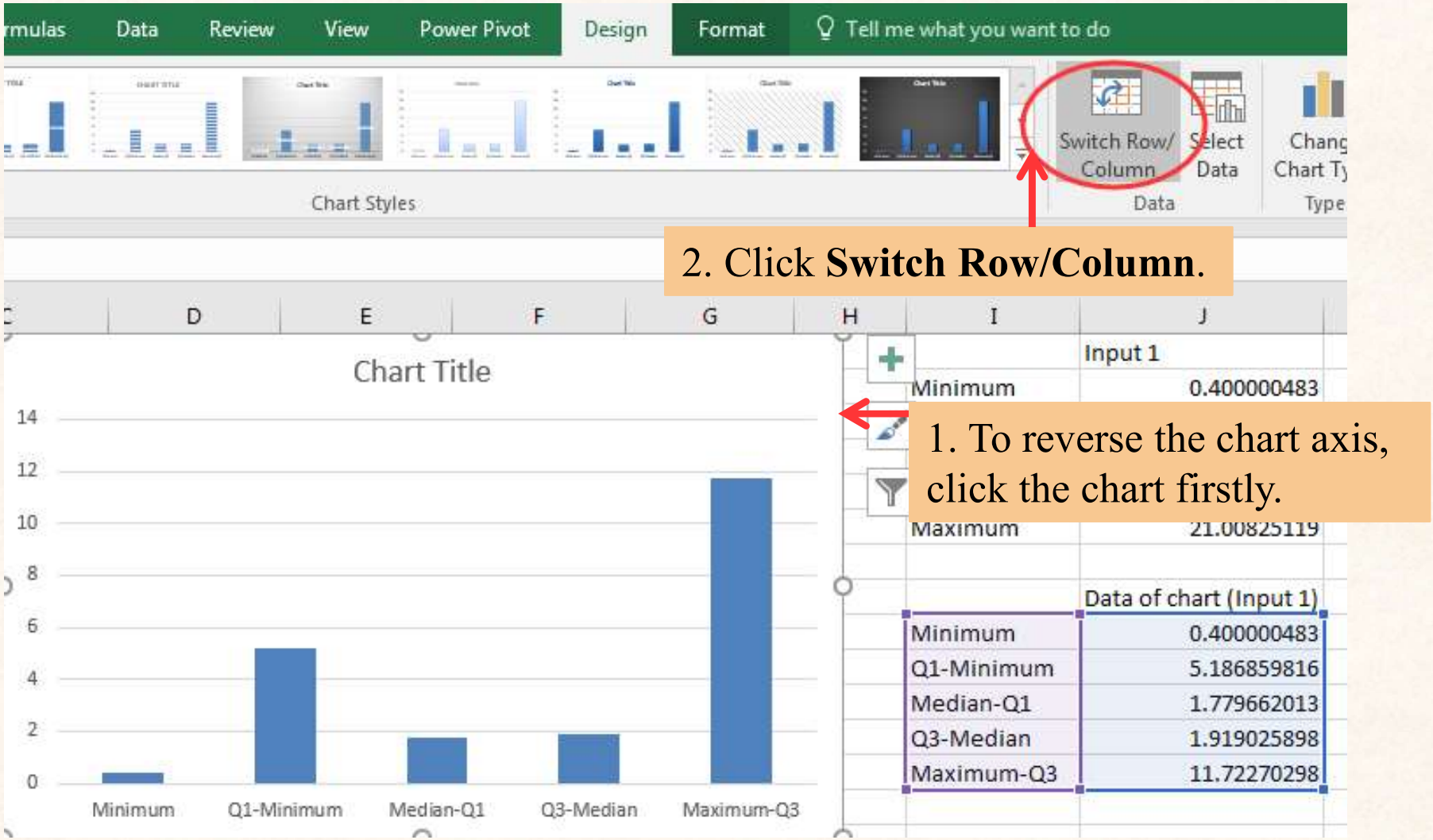
Part B.2 Box Plot (Box-whisker Plot)

- Step 3: Create a stacked column chart



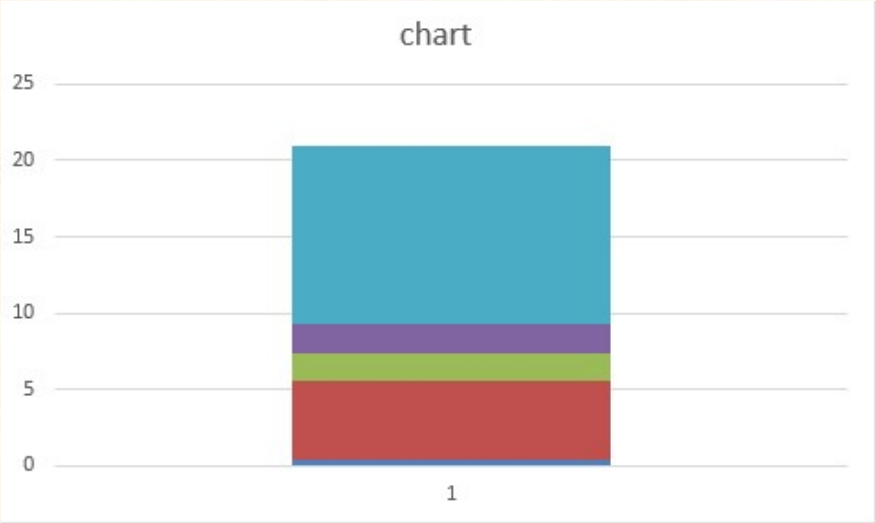
Part B.2 Box Plot (Box-whisker Plot)

- Step 3: Create a stacked column chart

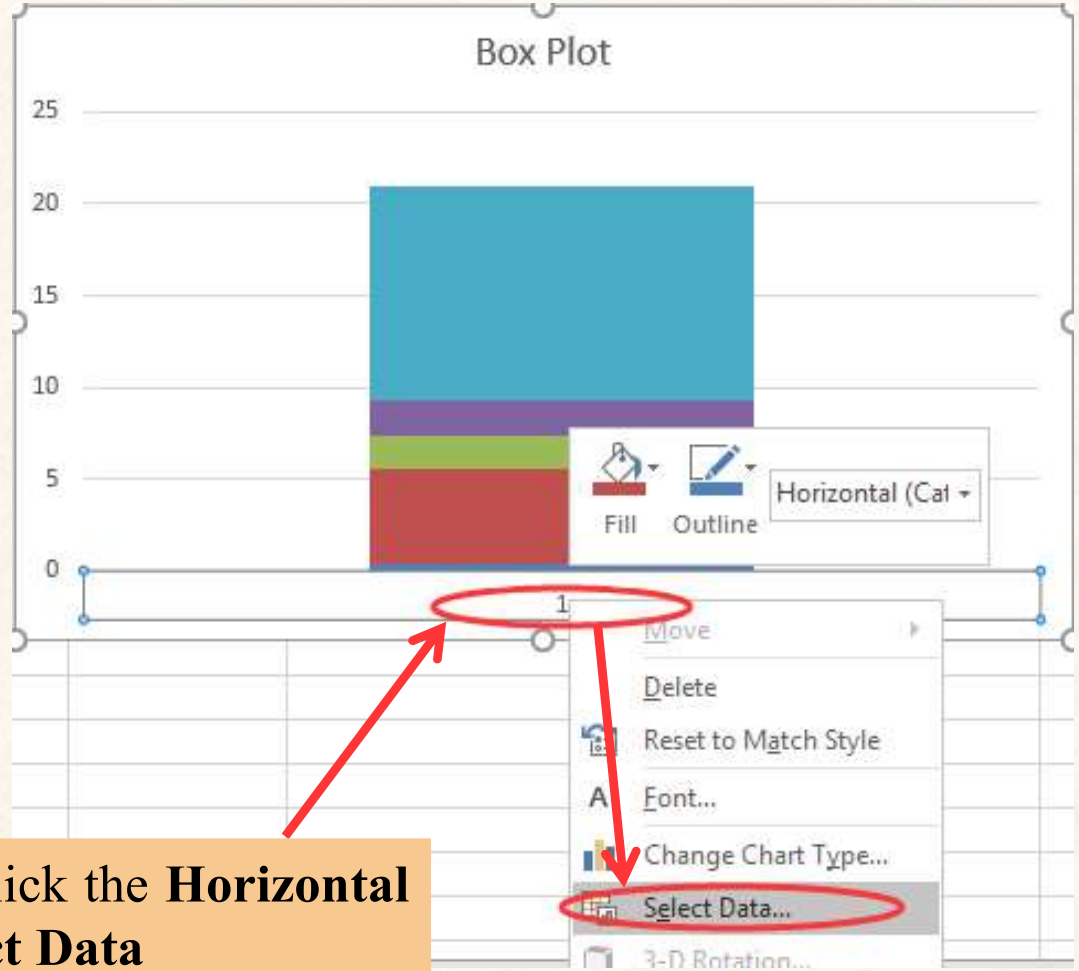


Part B.2 Box Plot (Box-whisker Plot)

- Step 3: Create a stacked column chart



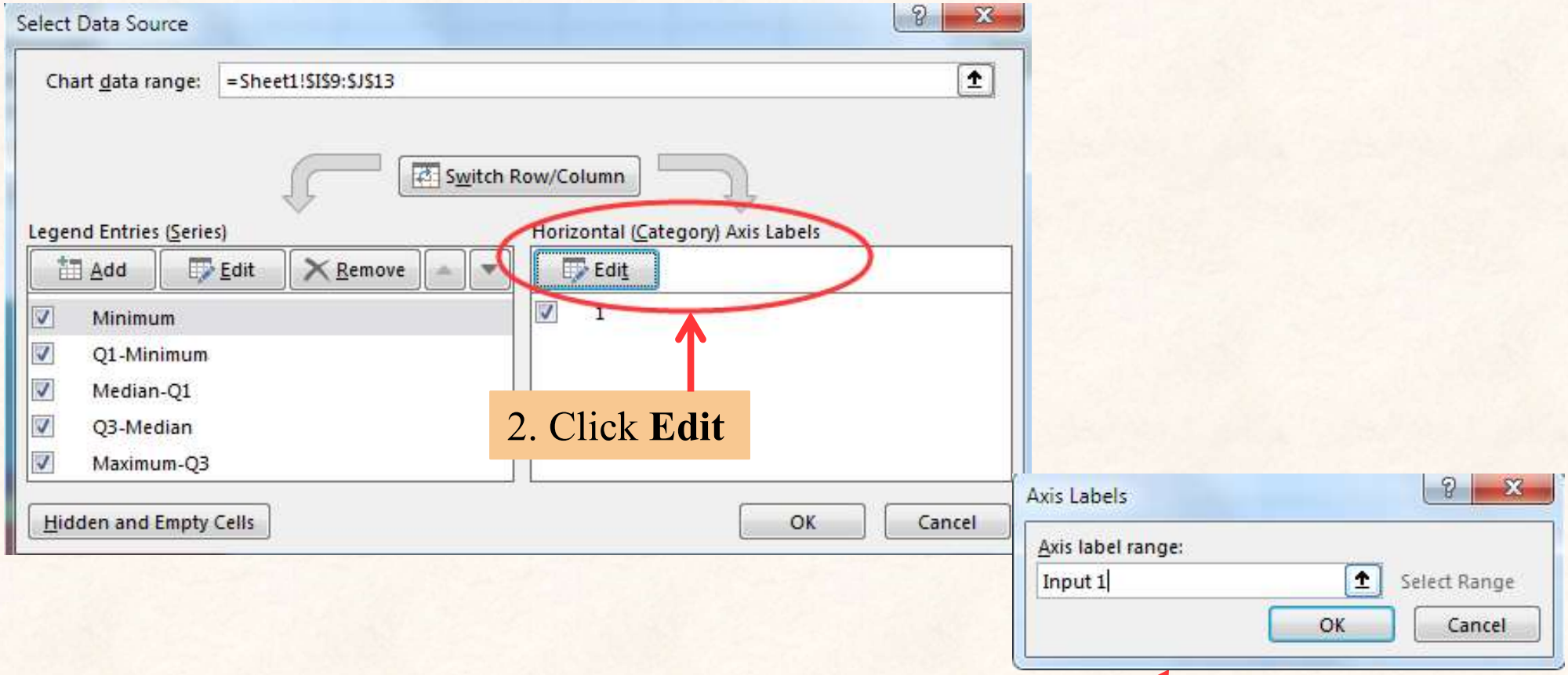
Original stacked column chart



1. To rename your columns, right click the **Horizontal (Category) axis labels** side -> **Select Data**

Part B.2 Box Plot (Box-whisker Plot)

- Step 3: Create a stacked column chart

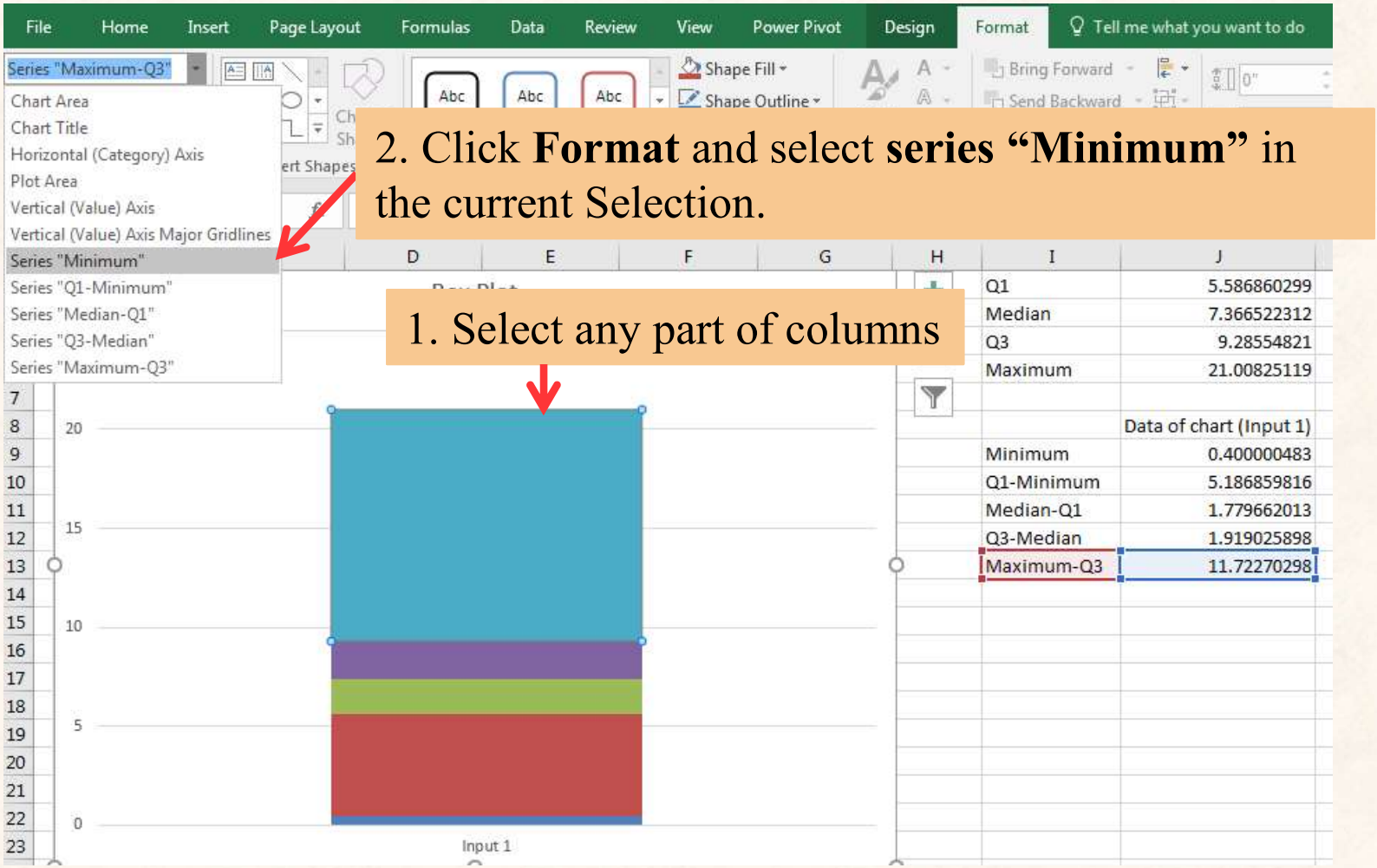


2. Click **Edit**

3. A new window pops out. Change the name as you want and click **OK**.

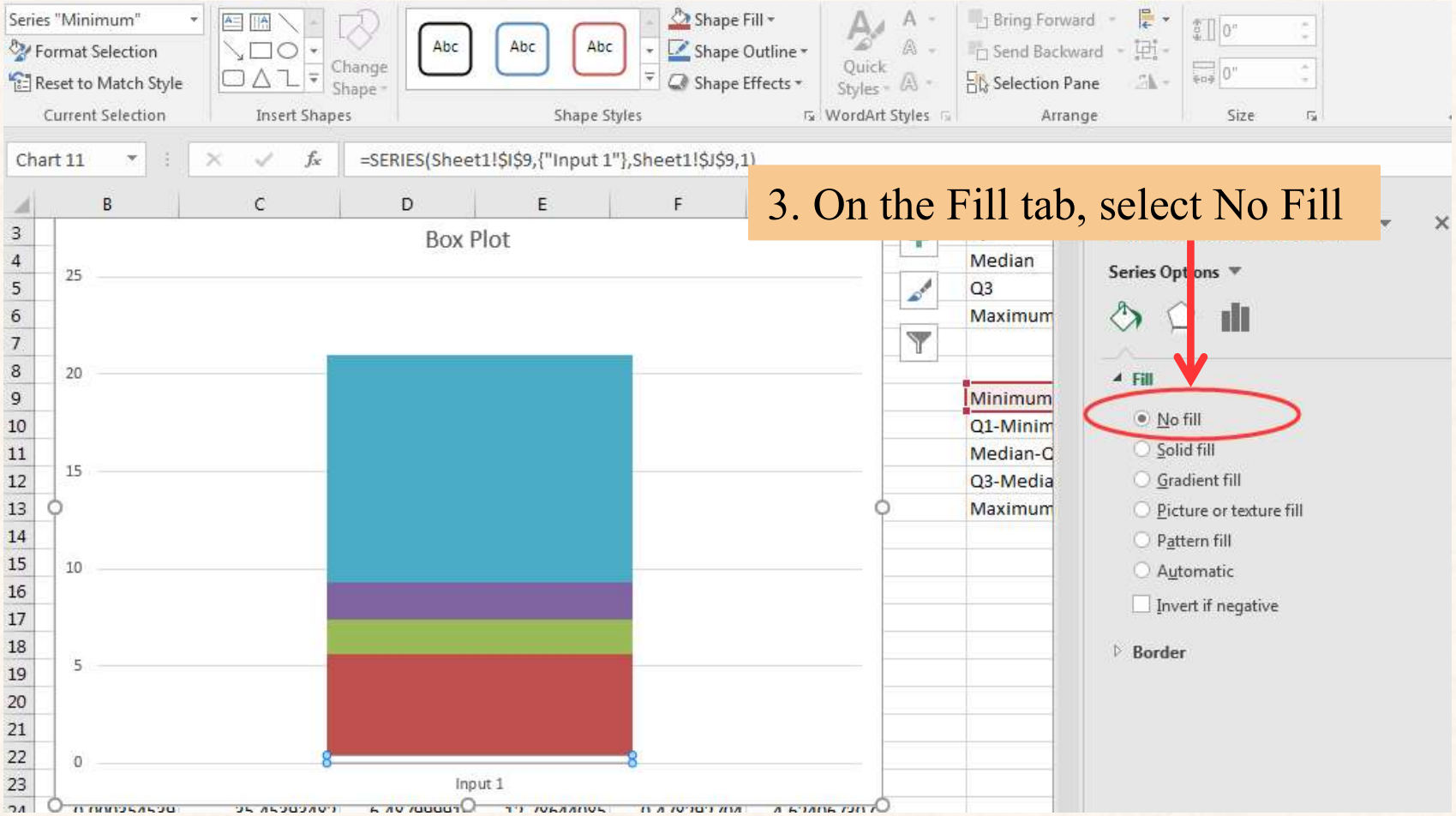
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.1 Hidden the bottom data series (minimum)



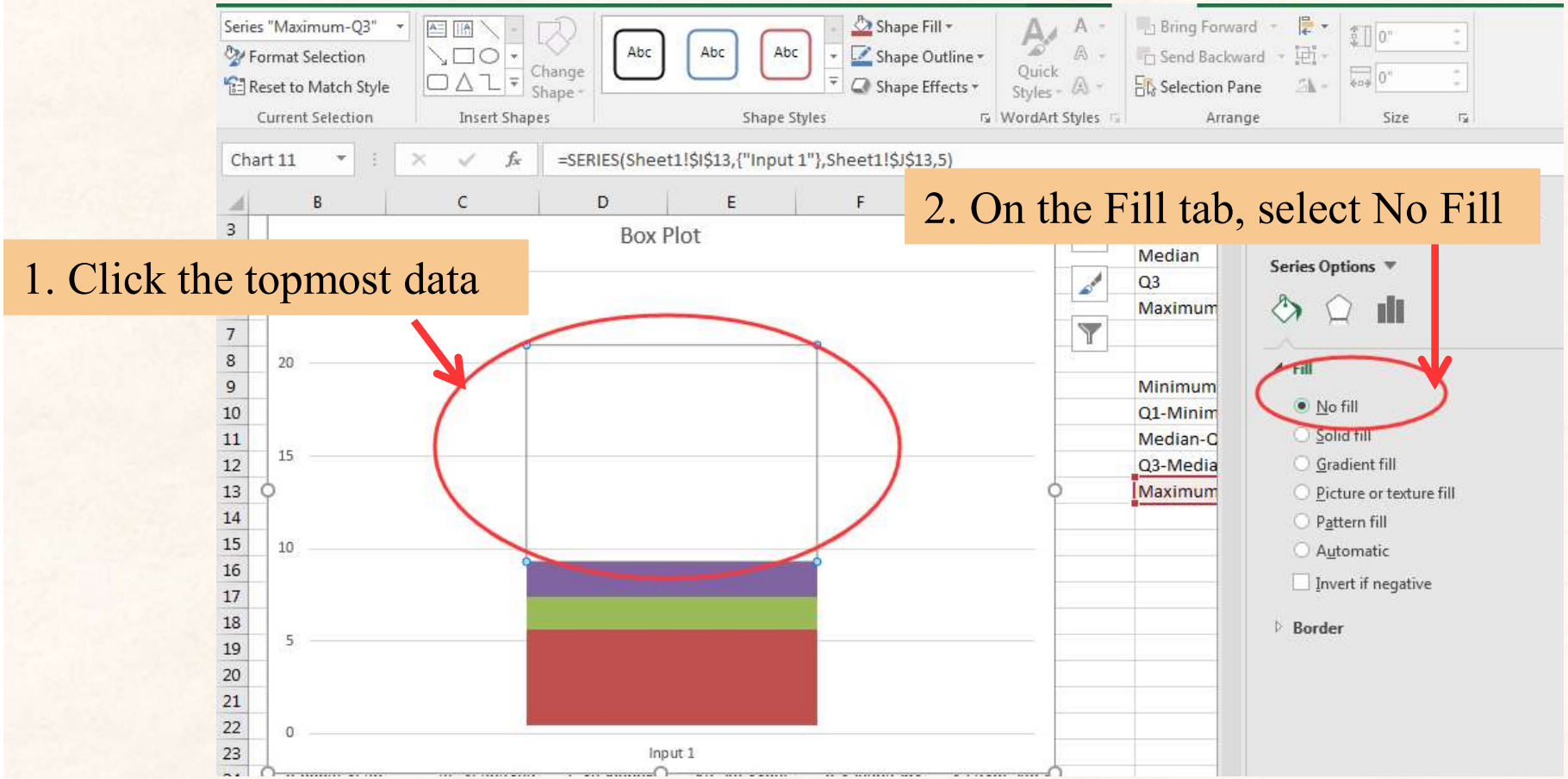
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.1 Hidden the bottom data series (minimum)



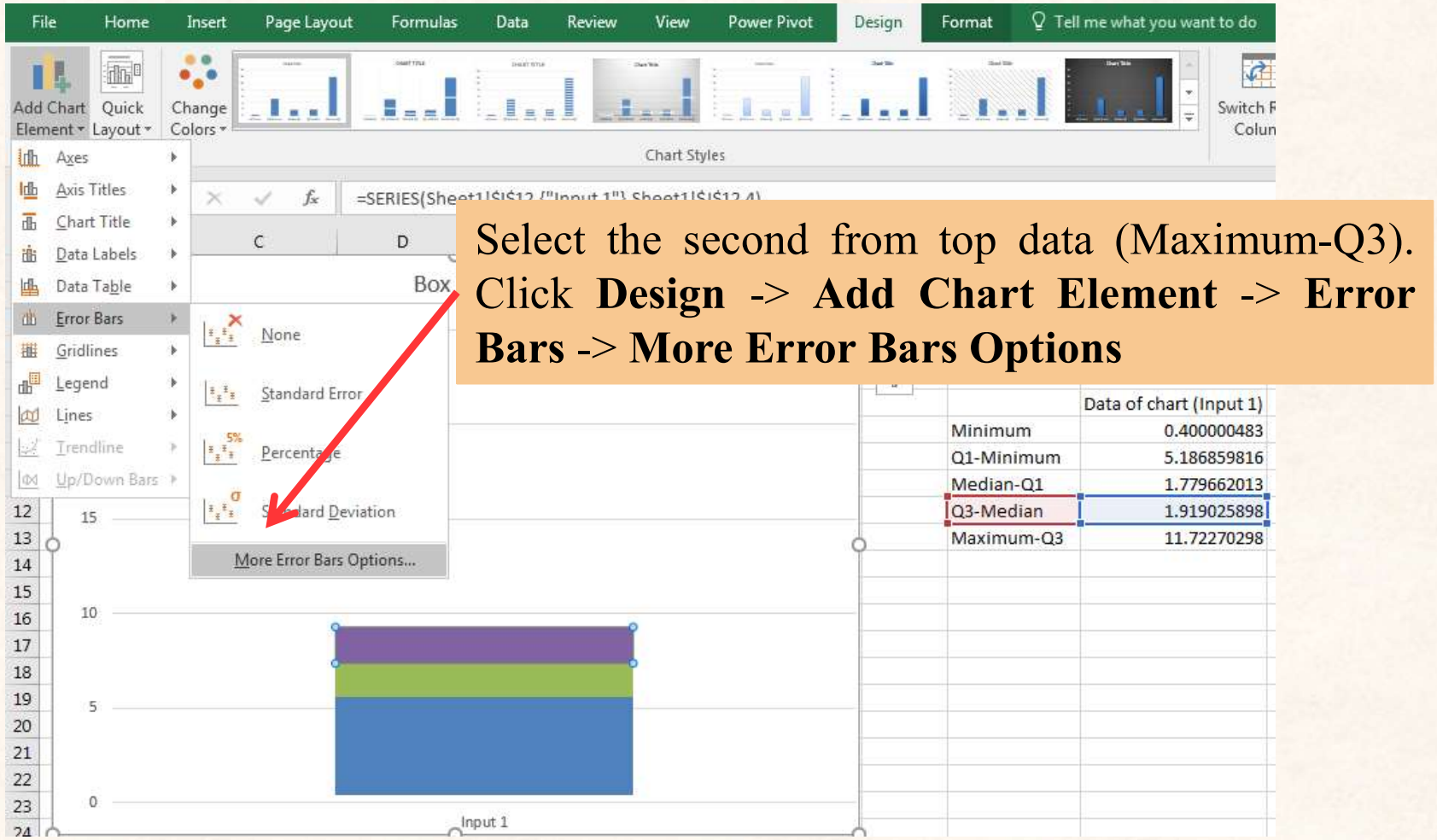
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.2 Create whiskers for the box plot



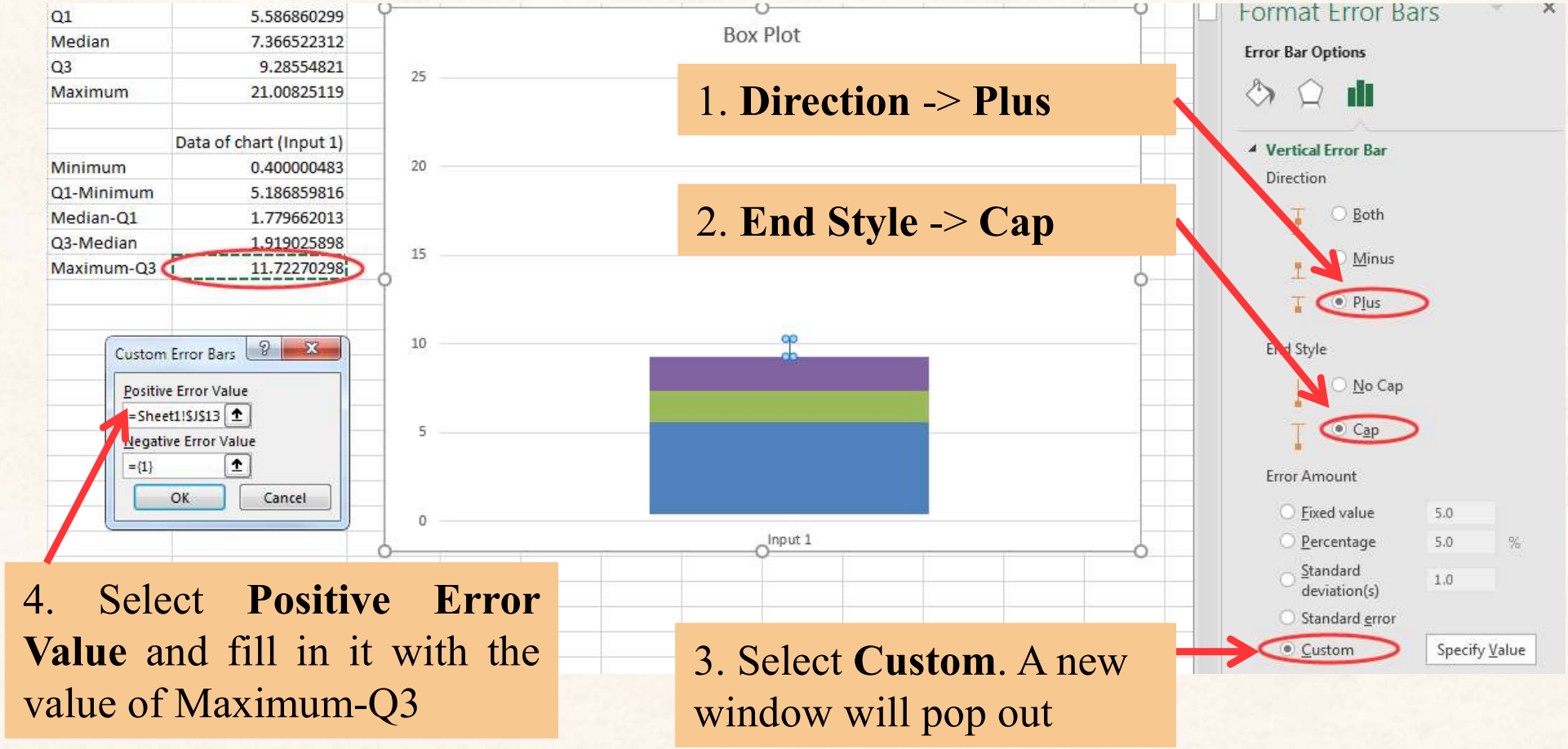
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.2 Create whiskers for the box plot



Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.2 Create whiskers for the box plot



1. **Direction -> Plus**

2. **End Style -> Cap**

3. Select **Custom**. A new window will pop out

4. Select **Positive Error Value** and fill in it with the value of Maximum-Q3

Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.2 Create whiskers for the box plot

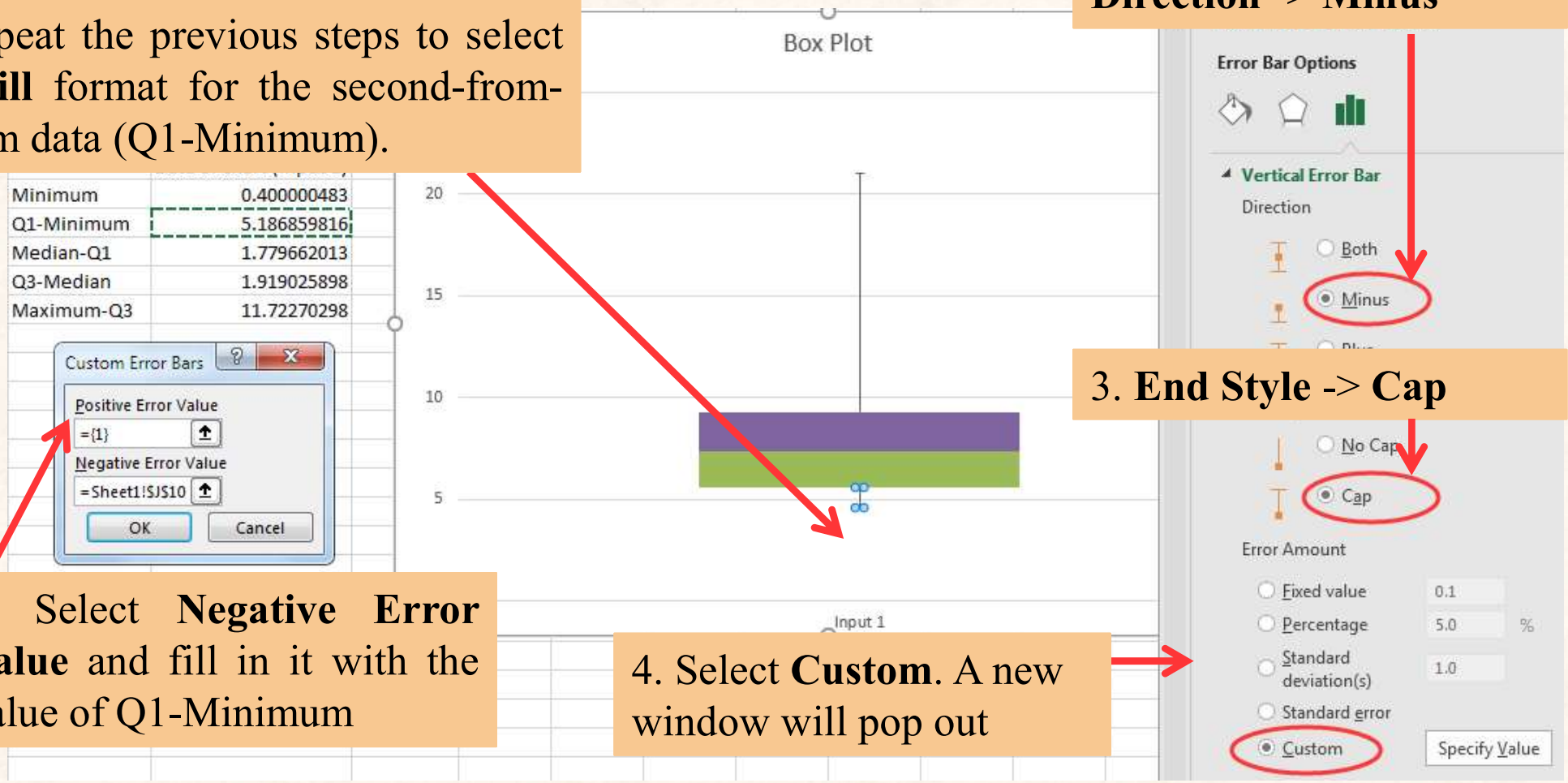
1. Repeat the previous steps to select **No Fill** format for the second-from-bottom data (Q1-Minimum).

2. In **Error Bar Options**, **Direction -> Minus**

3. **End Style -> Cap**

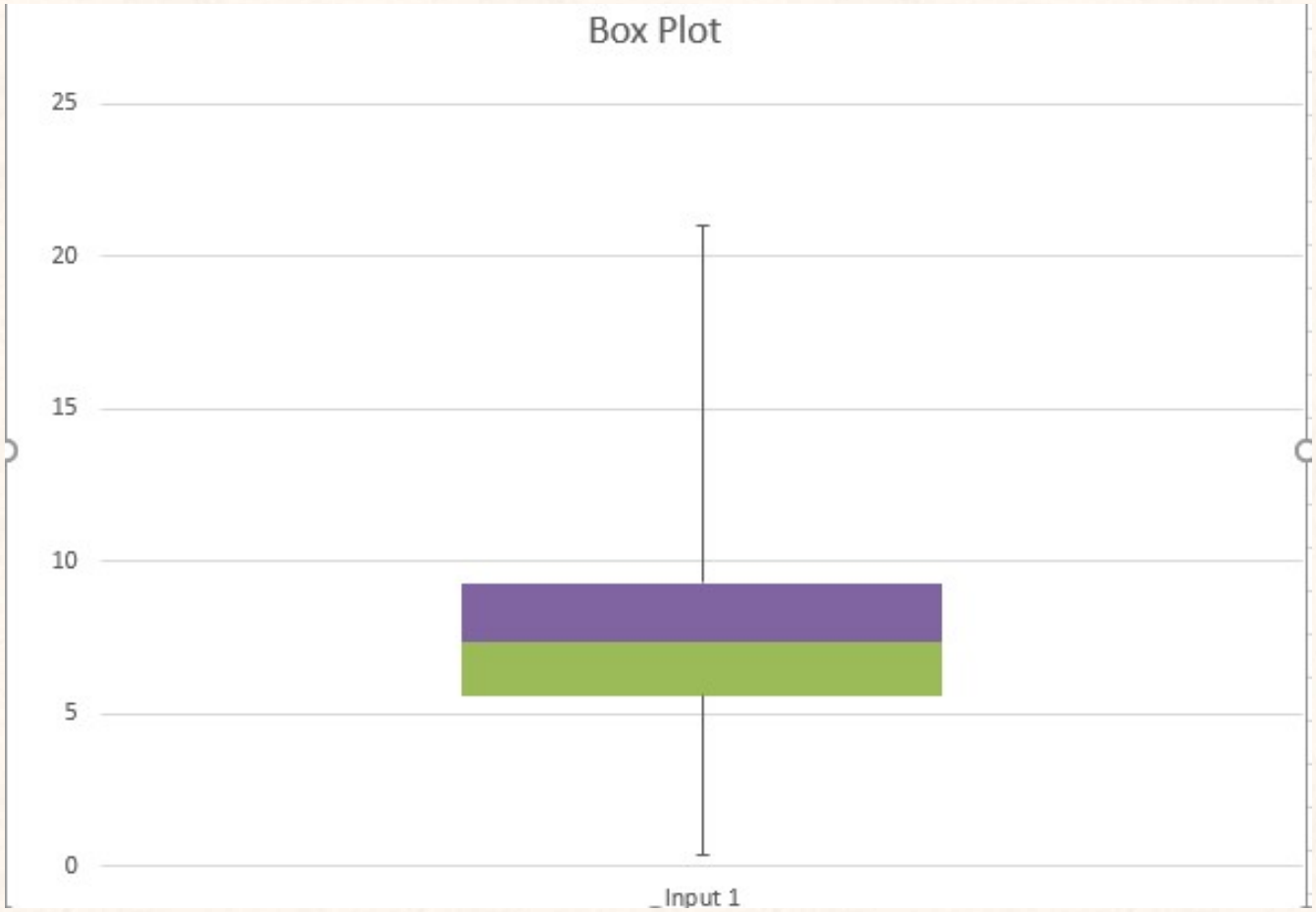
4. Select **Custom**. A new window will pop out

5. Select **Negative Error Value** and fill in it with the value of Q1-Minimum



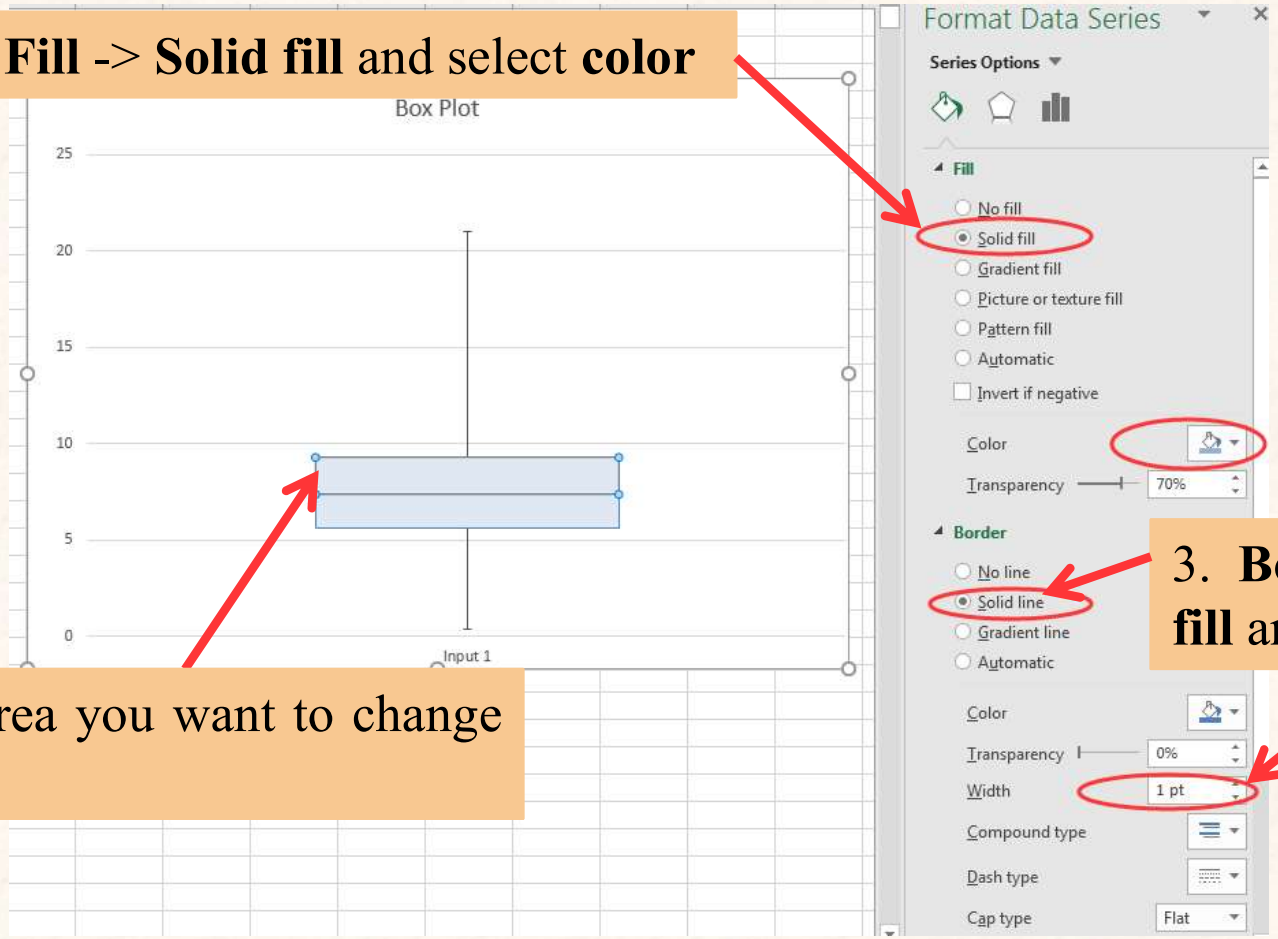
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.2 Create whiskers for the box plot



Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
  - 4.3 color the middle area. Box plots are usually drawn in one fill color, with a slight outline border.



2. **Fill** -> **Solid fill** and select **color**

1. Select the area you want to change its color.

3. **Border** -> **Solid fill** and select **width**

Exercise

- 1. Calculate descriptive statistics of Input2, Input 3 and Response
- 2. Plot histograms and box plots for Input 2, Input 3 and Response. You can try to plot box- whiskers of Input 1, Input 2, Input 3 and response in the same figure.