V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 1

Create Sampling Distributions from a Single Die in Excel 2013

Milo Schield

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

Slides at: www.StatLit.org/pdf/ Excel2013-Sampling-1Die-Demo-Slides.pdf

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 2

The Goal and Approach

Goal: to create the sampling distribution from a single die with various sample sizes using Excel 2013.

A preformatted data spreadsheet is at www.StatLit.org/xls/Excel2013-Sampling-1Die-Data.xlsx

This step-by-step demo is at www.StatLit.org/pdf/
- Excel2013-Sampling-1Die-Demo-Slides.pdf

A picture of the output is at www.StatLit.org/pdf/ Excel2013-Sampling-1Die-Demo-Output.pdf

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 3

Steps in Creating Sampling Distributions for a single die

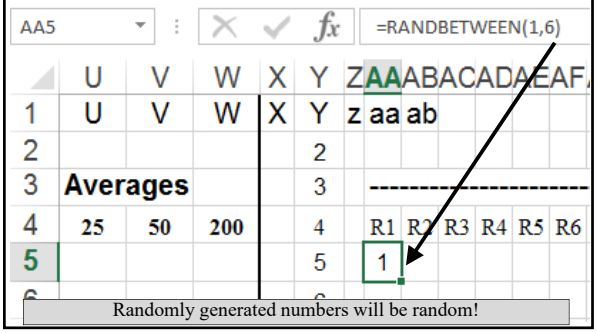
Access Pre-formatted Data Worksheet:

1. Insert RandBetween(1, 6) to simulate throw of die.
2. Create row averages: samples of 4, 16, 25, 50, 200.
3. Calculate population statistics for a single die.
4. Calculate summary sample statistics by sample size.
5. Group row averages into frequency bins.
6. Create line-graph histograms by sample size.

Upload completed worksheet.

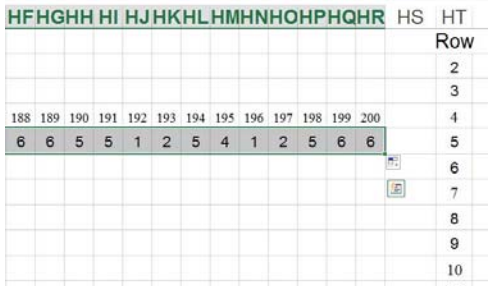
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 4

1a: Insert RandBetween(1,6) Function in AA5



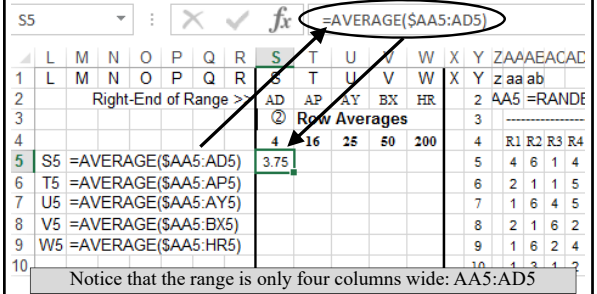
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 5

1b: Drag over 200 columns to the right to HR5



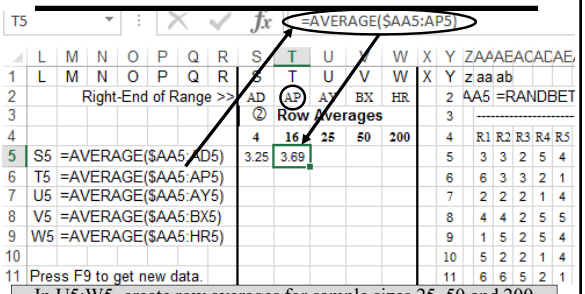
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 6

2a: In S5, create row average with sample size four



V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 7

2b: In T5, create row average with sample size of 16

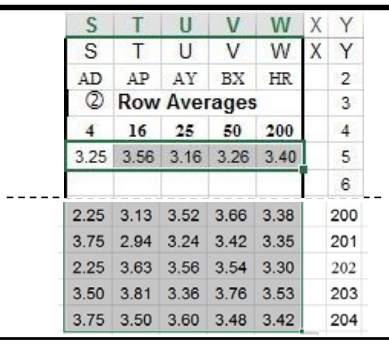


In U5:W5, create row averages for sample sizes 25, 50 and 200

Right-end of ranges are shown in S2:W2

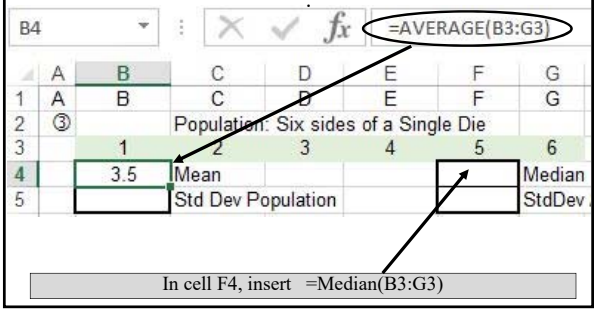
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 8

2d: Drag First Row Averages, S5:W5, down to Row 204.



V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 9

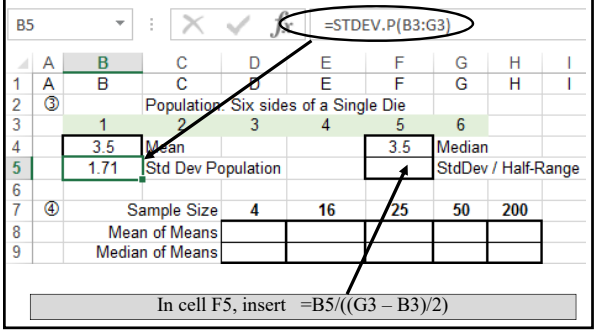
3a: Generate mean (average) [and median] for a six-sided die



In cell F4, insert =Median(B3:G3)

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 10

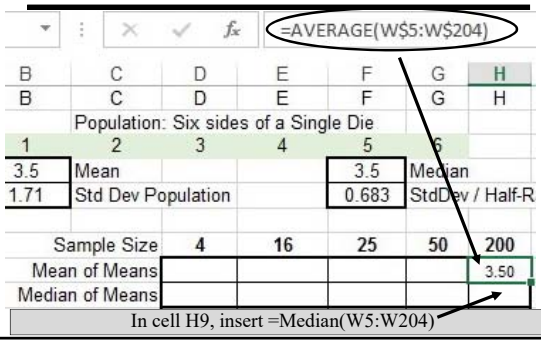
3b: Generate population std. deviation for a fair die



In cell F5, insert =B5/((G3 – B3)/2)

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 11

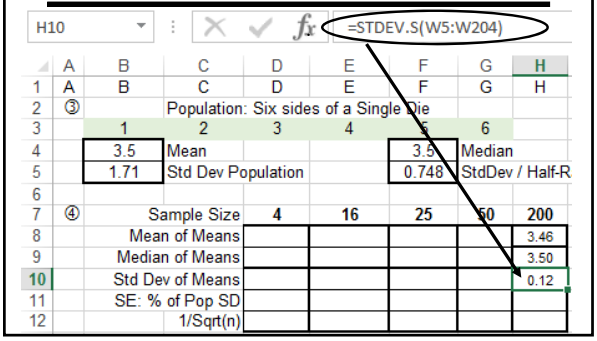
4a: Generate Mean [& Median] of Row Means for Sample Size 200



In cell H9, insert =Median(W5:W204)

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 12

4b: Generate Sample Std.Dev of Row Means for Size 200



V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 13

4c: H11 Ratio of Std. Error to Population Std. Deviation

H11 =H10/$B5

	A	B	C	D	E	F	G	H
1	A	B	C	D	E	F	G	H
2	③		Population: Six sides of a Single Die					
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Population			0.748	StdDev / Half-R	
6								
7	④	Sample Size	4	16	25	50	200	
8		Mean of Means					3.47	
9		Median of Means					3.50	
10		Std Dev of Means					0.12	
11		SE: % of Pop SD					7%	
12		1/Sqrt(n)						

Insert =1/SQRT(H7) into cell H12

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 14

4d: Drag H8:H12 column left-ward to fill out the table

D11 =D10/$B5

	A	B	C	D	E	F	G	H
1	A	B	C	D	E	F	G	H
2	③		Population: Six sides of a Single Die					
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Population			0.748	StdDev / Half-R	
6								
7	④	Sample Size	4	16	25	50	200	
8		Mean of Means	3.52	3.51	3.52	3.53	3.52	
9		Median of Means	3.50	3.53	3.52	3.52	3.51	
10		Std Dev of Means	0.85	0.41	0.34	0.24	0.12	
11		SE: % of Pop SD	50%	24%	20%	14%	7%	
12		1/Sqrt(n)	50%	25%	20%	14%	7%	

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 15

5a: Insert COUNTIF function in H17

H17 =COUNTIF(W$5:W$204, "<="&$C17)

	A	B	C	D	E	F	G	H	I
13									
14	⑤		Rolling a Single Six-Sided Die						
15			------ Distribution of Averages by Sample size ----						
16		Midpoint	Max	4	16	25	50	200	
17		2.10	2.00					0	
18		2.10	2.20						
19		2.30	2.40						
20		2.50	2.60						
21		2.70	2.80						
22		2.90	3.00						
23		3.10	3.20						
24		3.30	3.40						

Insert $ sign in before Column in range; before Row in single cell!

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 16

5b: Insert "=CountIF() –Sum()" function in 2nd row: H18

=COUNTIF(W$5:W$204,"<="&$C18)-SUM(H$17:H17)

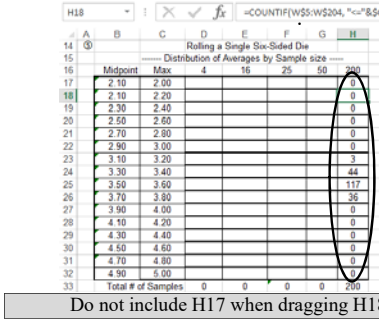
	C	D	E	F	G	H	I	J
	Rolling a Single Six-Sided Die							
	------ Distribution of Averages by Sample size ----							
	Max	4	16	25	50	200		
	2.00					0		
	2.20					0		
	2.40							
	2.60							
	2.80							
	3.00							
	3.20							
	3.40							
	3.60							

H18: Equivalent to =COUNTIF(W$5:W$204, "<="&$C18) - COUNTIF(W$5:W$204, "<="&$C17)

Insert single $ sign in Sum function before first row.

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 17

5c: Drag H18, =CountIF() –Sum(), formula down to bottom row



Midpoint	Max	4	16	25	50	200
2.10	2.00					0
2.10	2.20					0
2.30	2.40					0
2.50	2.60					0
2.70	2.80					0
2.90	3.00					0
3.10	3.20					3
3.30	3.40					44
3.50	3.60					117
3.70	3.80					36
3.90	4.00					0
4.10	4.20					0
4.30	4.40					0
4.50	4.60					0
4.70	4.80					0
4.90	5.00					0
Total # of Samples		0	0	0	0	200

Drag entire right column (H17:H32) left-ward to fill out frequency table.

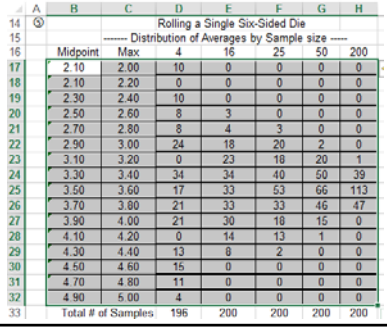
Insert SUM at bottom of each column.

May be different from 200.

Do not include H17 when dragging H18 downward!!!

V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 18

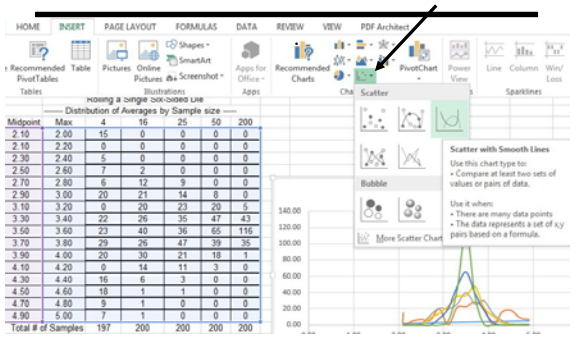
6a: Select Data (B17:H32) to use in Histogram



Midpoint	Max	4	16	25	50	200
2.10	2.00	10	0	0	0	0
2.10	2.20	0	0	0	0	0
2.30	2.40	10	0	0	0	0
2.50	2.60	8	3	0	0	0
2.70	2.80	8	4	3	0	0
2.90	3.00	24	18	20	2	0
3.10	3.20	0	23	18	20	1
3.30	3.40	34	34	40	50	39
3.50	3.60	17	33	53	66	113
3.70	3.80	21	33	33	46	47
3.90	4.00	21	30	18	15	0
4.10	4.20	0	14	13	1	0
4.30	4.40	13	8	2	0	0
4.50	4.60	15	0	0	0	0
4.70	4.80	11	0	0	0	0
4.90	5.00	4	0	0	0	0
Total # of Samples		196	200	200	200	200

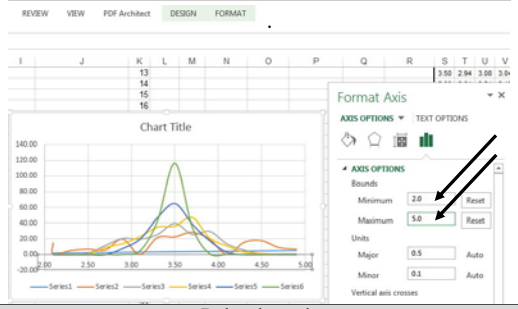
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 19

6b: Insert X-Y Scatter Plot



V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 20

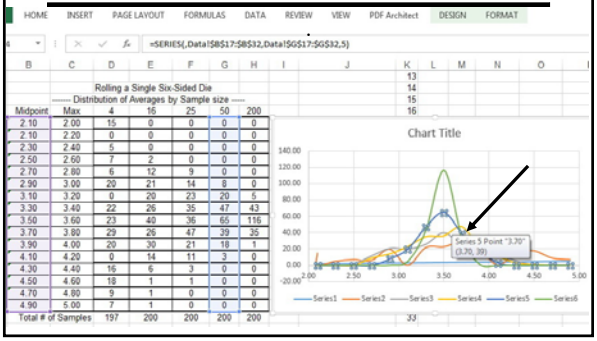
6c: Format Horizontal Axis: Change Min and Max



Delete legend.

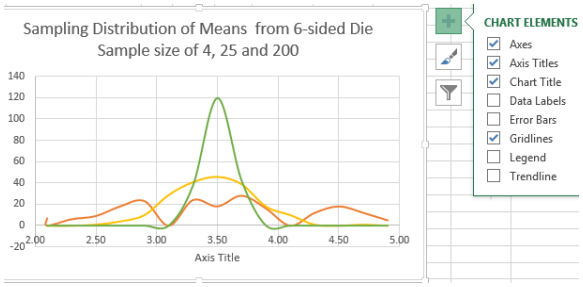
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 21

6d: Select/delete Series 5 (size 50), 3 (size 16) and 1 (Max).



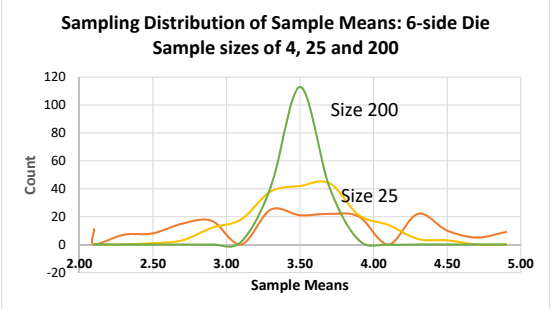
V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 22

6e: Insert Title and Horizontal Axis Text



V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 23

6f: Insert legends in text boxes. Final Result



V0G Create Sampling Distribution of a Single Die using COUNTIF in Excel 2013 24

Summary

When sampling from a process, the population "size" is "infinite". That doesn't influence the standard deviation.

Notice as sample size increases, the standard error (the std. deviation of the sample means) quickly decreases – as a percentage of the population standard deviation.

A sample of size 4 is expected to have a standard error that is only a half of the population standard deviation: a sample of 25 has a fifth, a sample of 100 has a tenth and a sample of 10,000 has a hundredth.

Create Sampling Distributions from a Single Die in Excel 2013

Milo Schield

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

Slides at: www.StatLit.org/pdf/Excel2013-Sampling-1Die-Demo-Slides.pdf

The Goal and Approach

Goal: to create the sampling distribution from a single die with various sample sizes using Excel 2013.

A preformatted data spreadsheet is at www.StatLit.org/xls/Excel2013-Sampling-1Die-Data.xlsx

This step-by-step demo is at www.StatLit.org/pdf/

- Excel2013-Sampling-1Die-Demo-Slides.pdf

A picture of the output is at www.StatLit.org/pdf/Excel2013-Sampling-1Die-Demo-Output.pdf

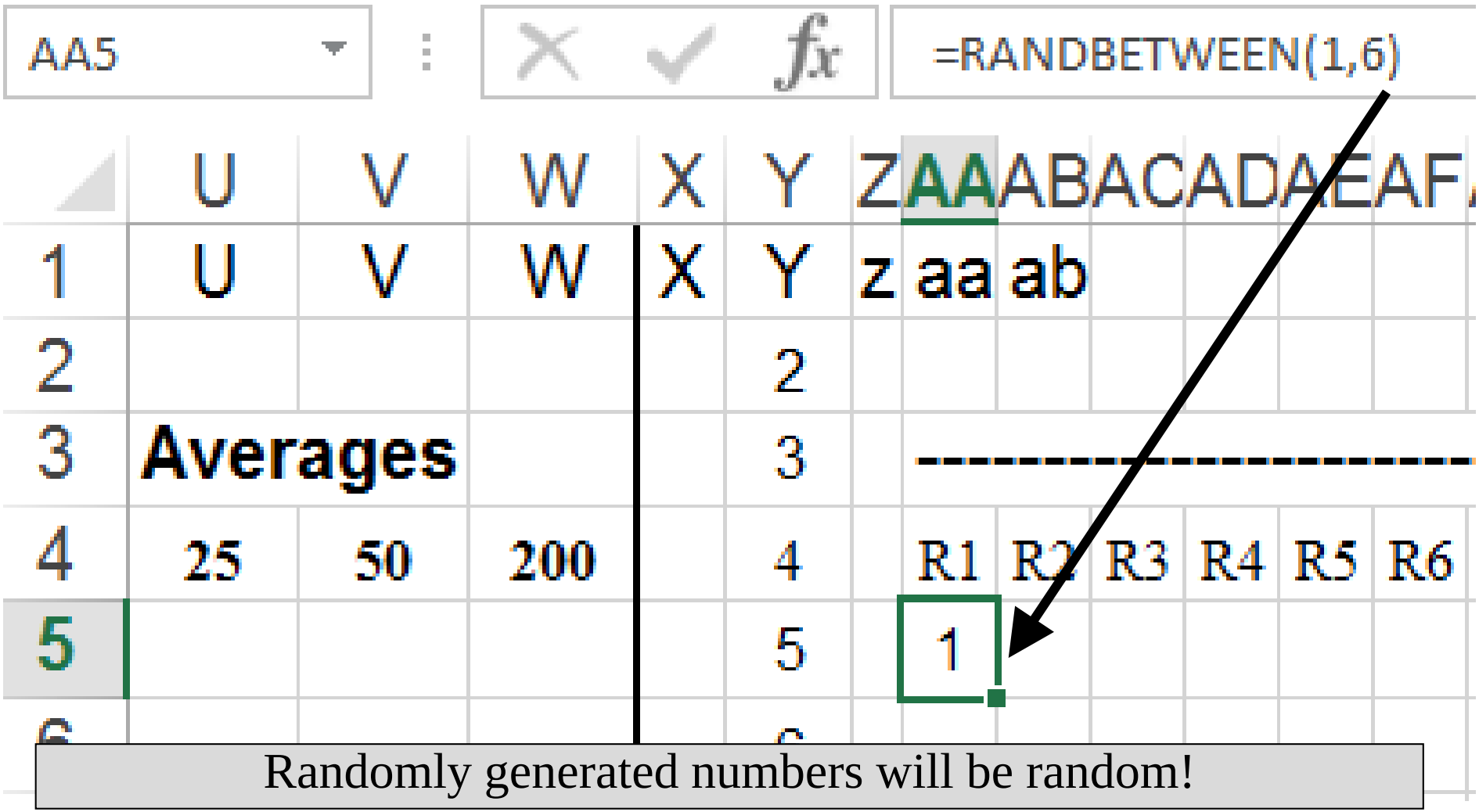
Steps in Creating Sampling Distributions for a single die

Access Pre-formatted Data Worksheet:

1. Insert RandBetween(1, 6) to simulate throw of die.
2. Create row averages: samples of 4, 16, 25, 50, 200.
3. Calculate population statistics for a single die.
4. Calculate summary sample statistics by sample size.
5. Group row averages into frequency bins.
6. Create line-graph histograms by sample size.

Upload completed worksheet.

1a: Insert RandBetween(1,6) Function in AA5



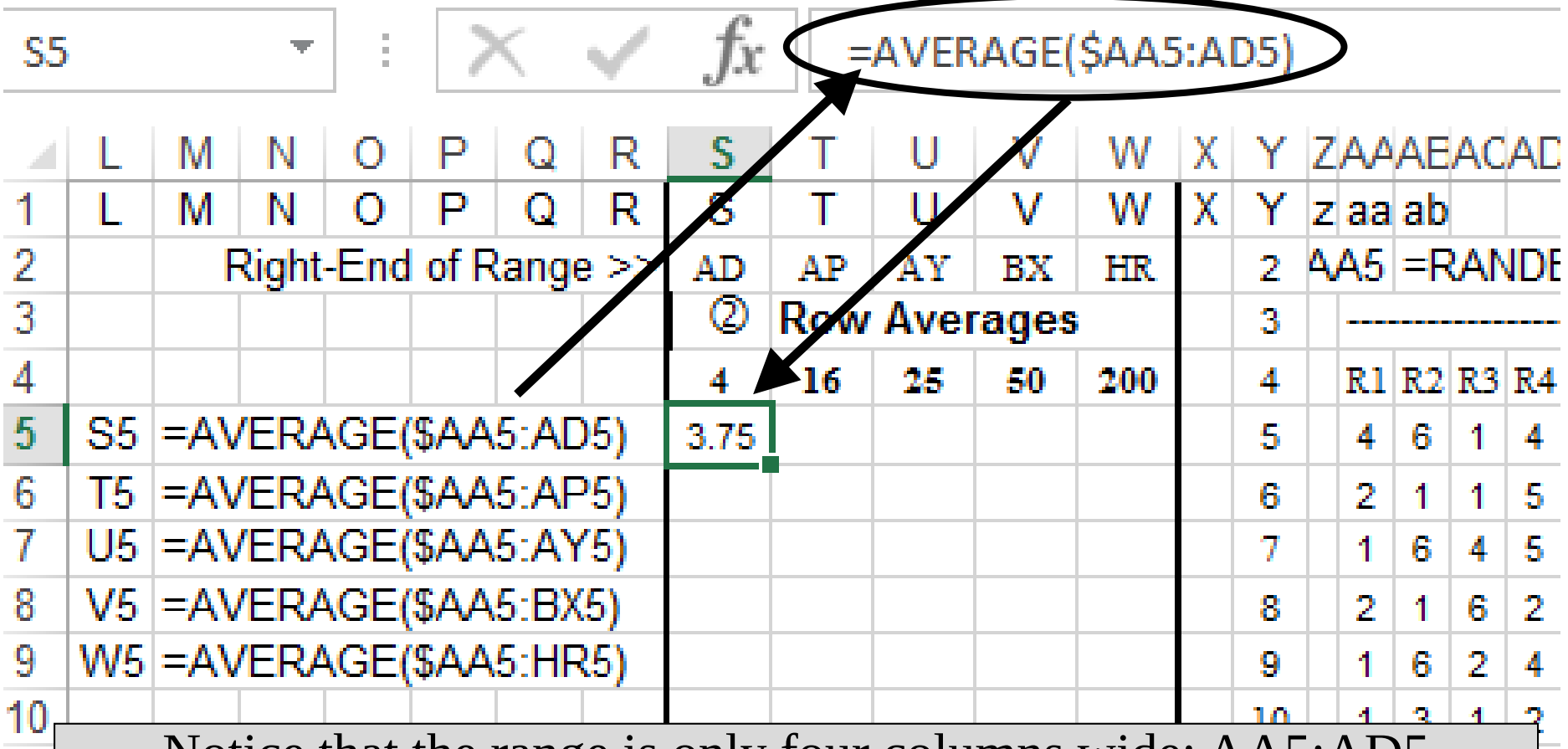
Randomly generated numbers will be random!

1b: Drag over 200 columns to the right to HR5

HF	HG	HH	HI	HJ	HK	HL	HM	HN	HO	HP	HQ	HR	HS	HT
														Row
														2
														3
188	189	190	191	192	193	194	195	196	197	198	199	200		4
6	6	5	5	1	2	5	4	1	2	5	6	6		5
														6
														7
														8
														9
														10

Drag this first row of random numbers down 200 rows to Row 204

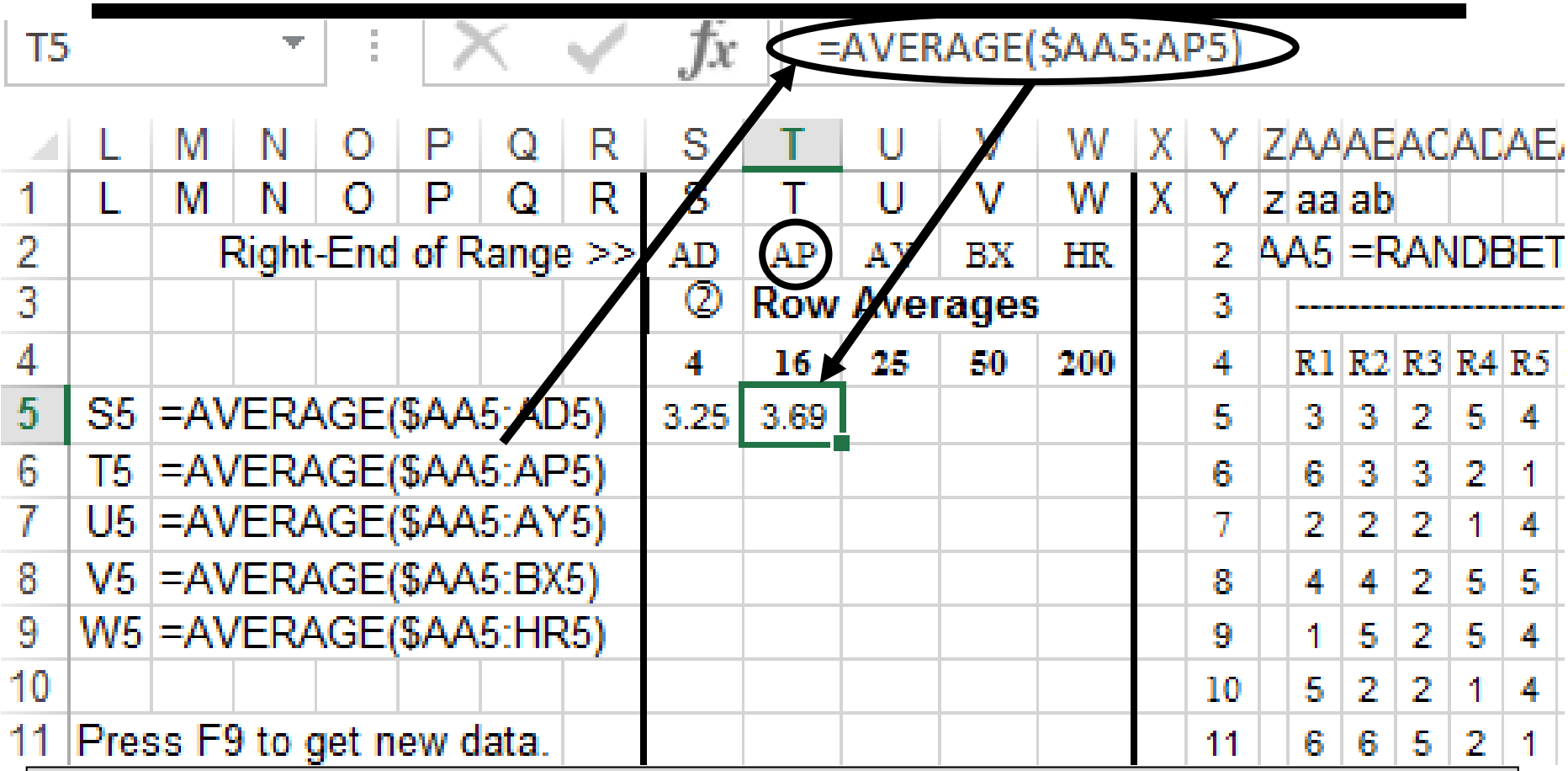
2a: In S5, create row average with sample size four



Notice that the range is only four columns wide: AA5:AD5

Averages of random numbers will be random!

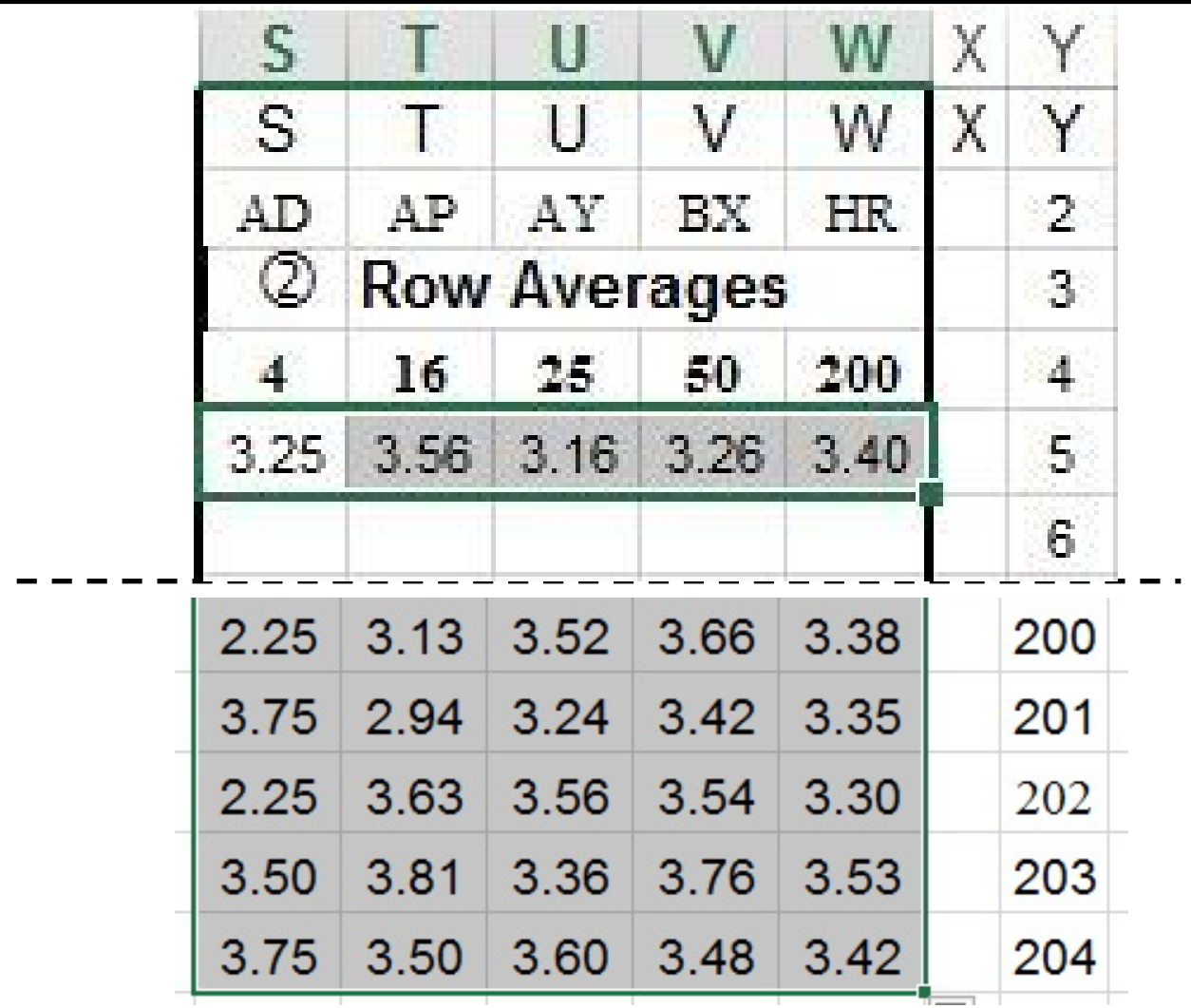
2b: In T5, create row average with sample size of 16



In U5:W5, create row averages for sample sizes 25, 50 and 200

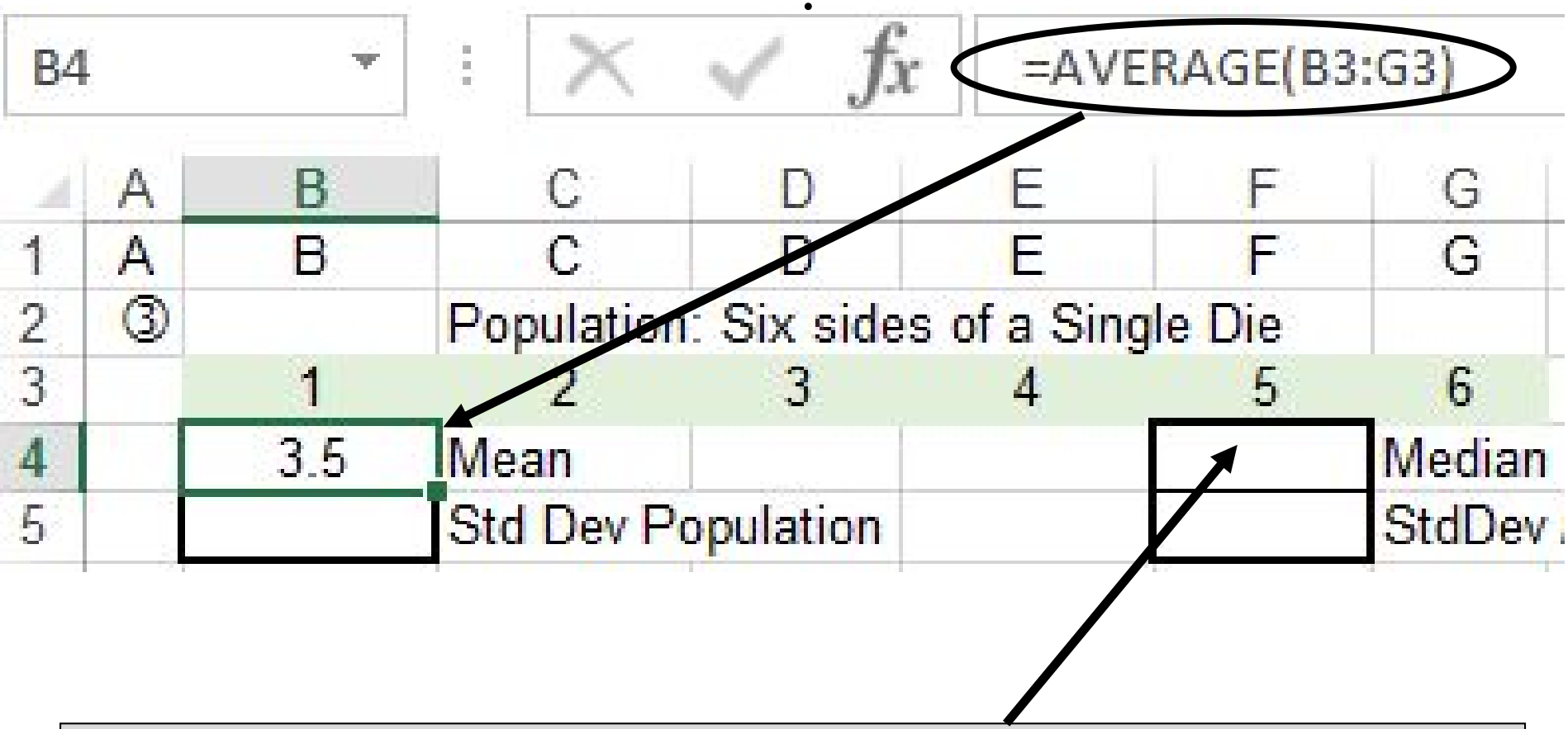
Right-end of ranges are shown in S2:W2

2d: Drag First Row Averages, S5:W5, down to Row 204.



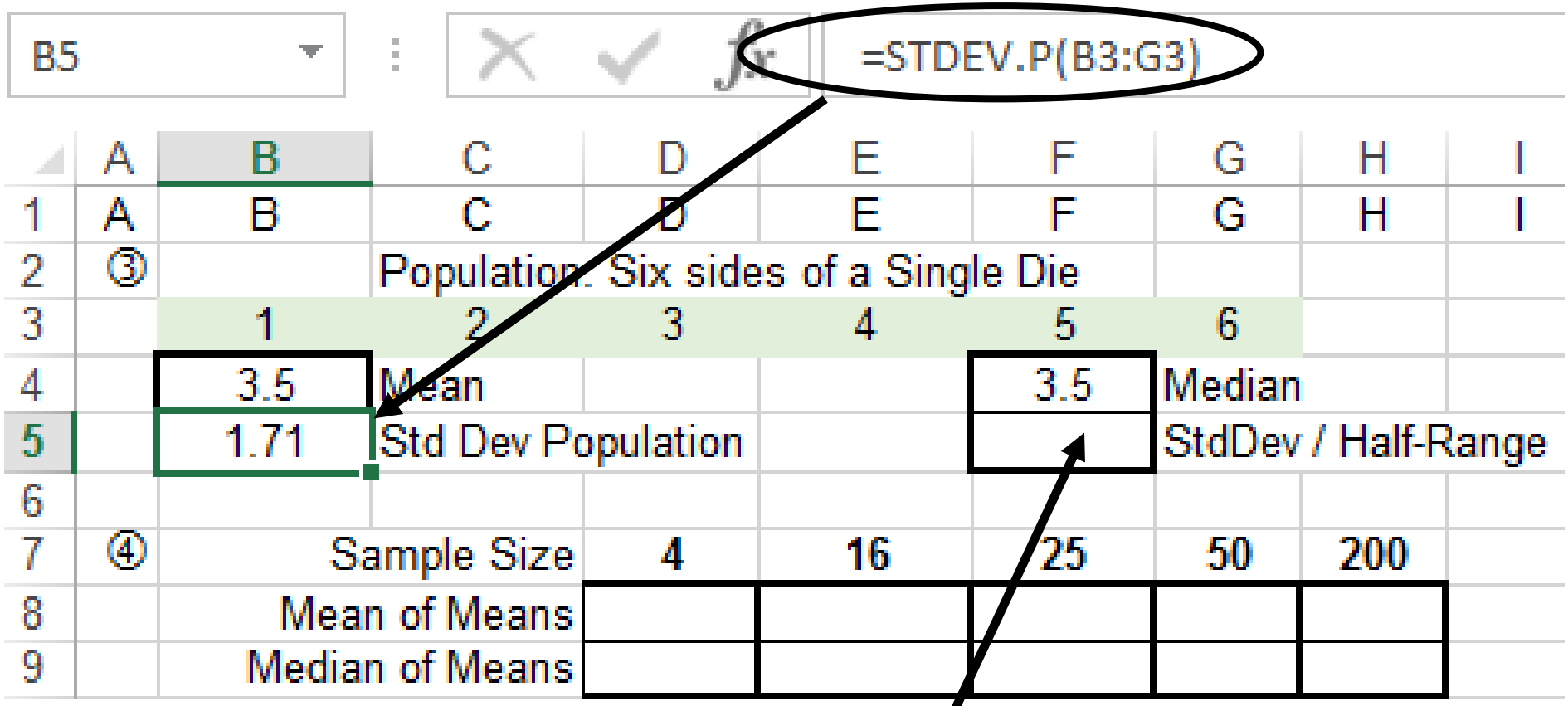
S	T	U	V	W	X	Y
S	T	U	V	W	X	Y
AD	AP	AY	BX	HR		2
②	Row Averages					3
4	16	25	50	200		4
3.25	3.56	3.16	3.26	3.40		5
						6
2.25	3.13	3.52	3.66	3.38		200
3.75	2.94	3.24	3.42	3.35		201
2.25	3.63	3.56	3.54	3.30		202
3.50	3.81	3.36	3.76	3.53		203
3.75	3.50	3.60	3.48	3.42		204

3a: Generate mean (average) [and median] for a six-sided die



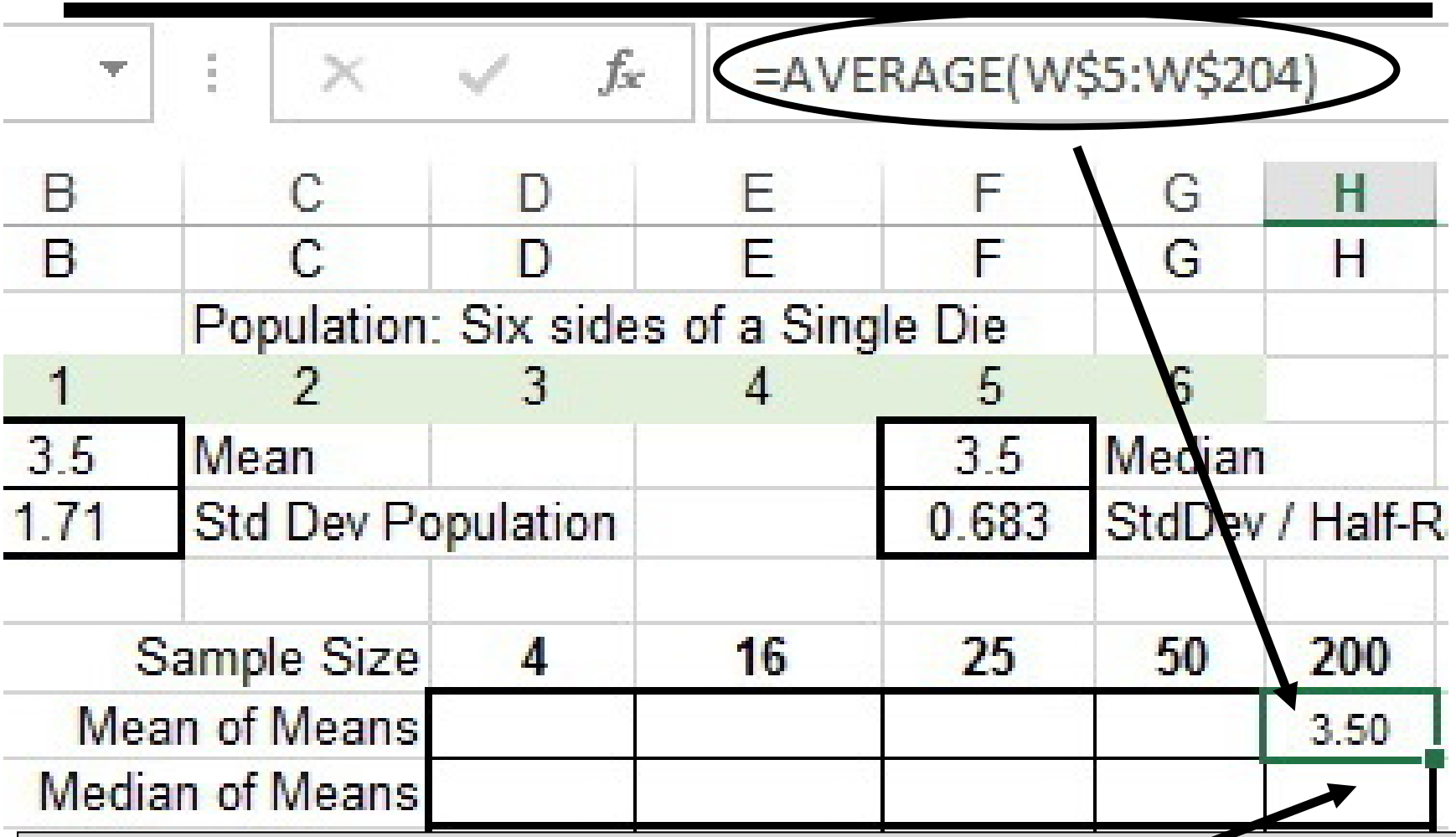
In cell F4, insert =Median(B3:G3)

3b: Generate population std. deviation for a fair die



In cell F5, insert =B5/((G3 – B3)/2)

4a: Generate Mean [& Median] of Row Means for Sample Size 200

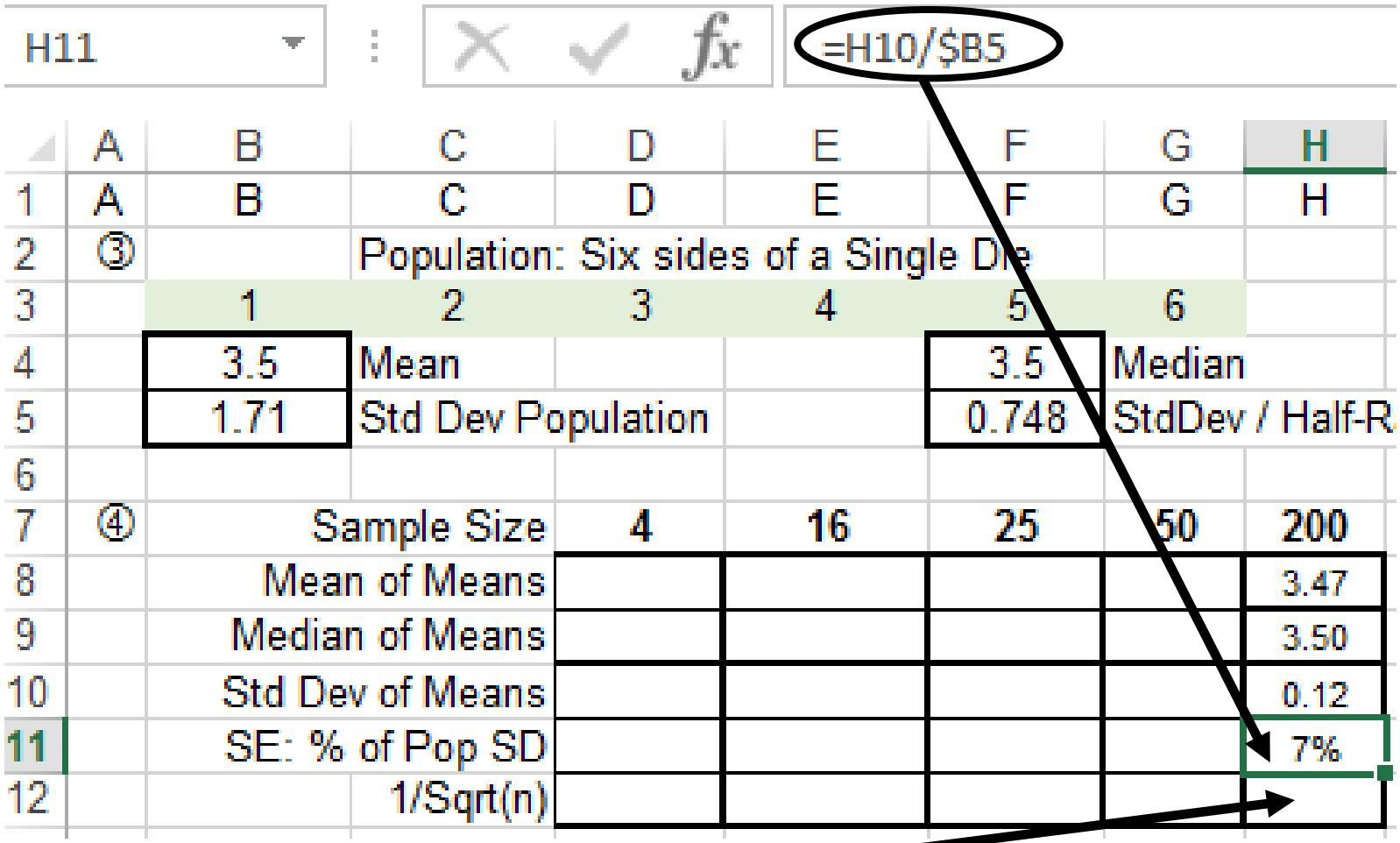


4b: Generate Sample Std.Dev of Row Means for Size 200

H10 =STDEV.S(W5:W204)

	A	B	C	D	E	F	G	H
1	A	B	C	D	E	F	G	H
2	③		Population: Six sides of a Single Die					
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Population			0.748	StdDev / Half-R	
6								
7	④	Sample Size	4	16	25	50	200	
8		Mean of Means					3.46	
9		Median of Means					3.50	
10		Std Dev of Means					0.12	
11		SE: % of Pop SD						
12		1/Sqrt(n)						

4c: H11 Ratio of Std. Error to Population Std. Deviation



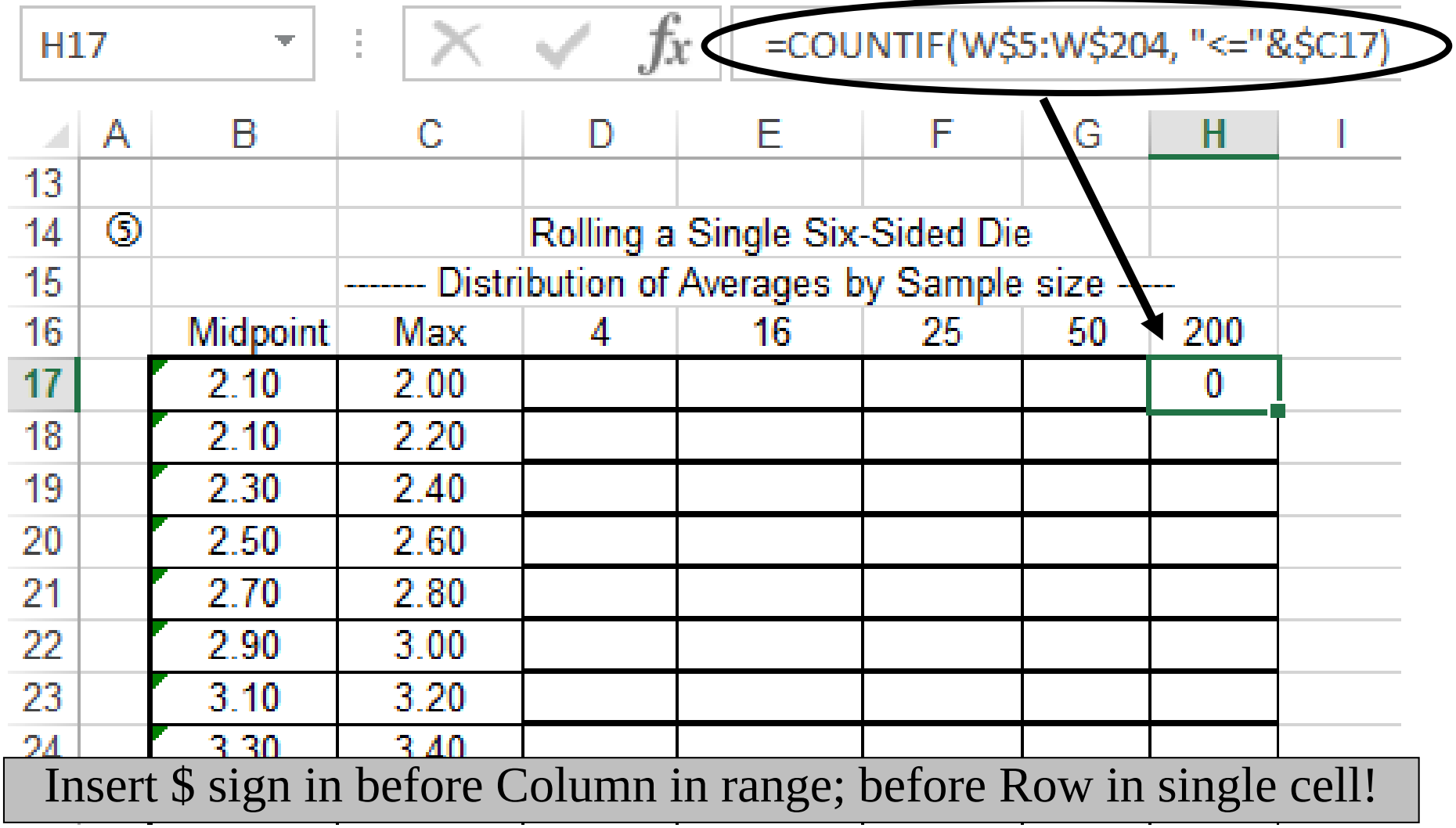
Insert =1/SQRT(H7) into cell H12

4d: Drag H8:H12 column left-ward to fill out the table

D11 =D10/$B5

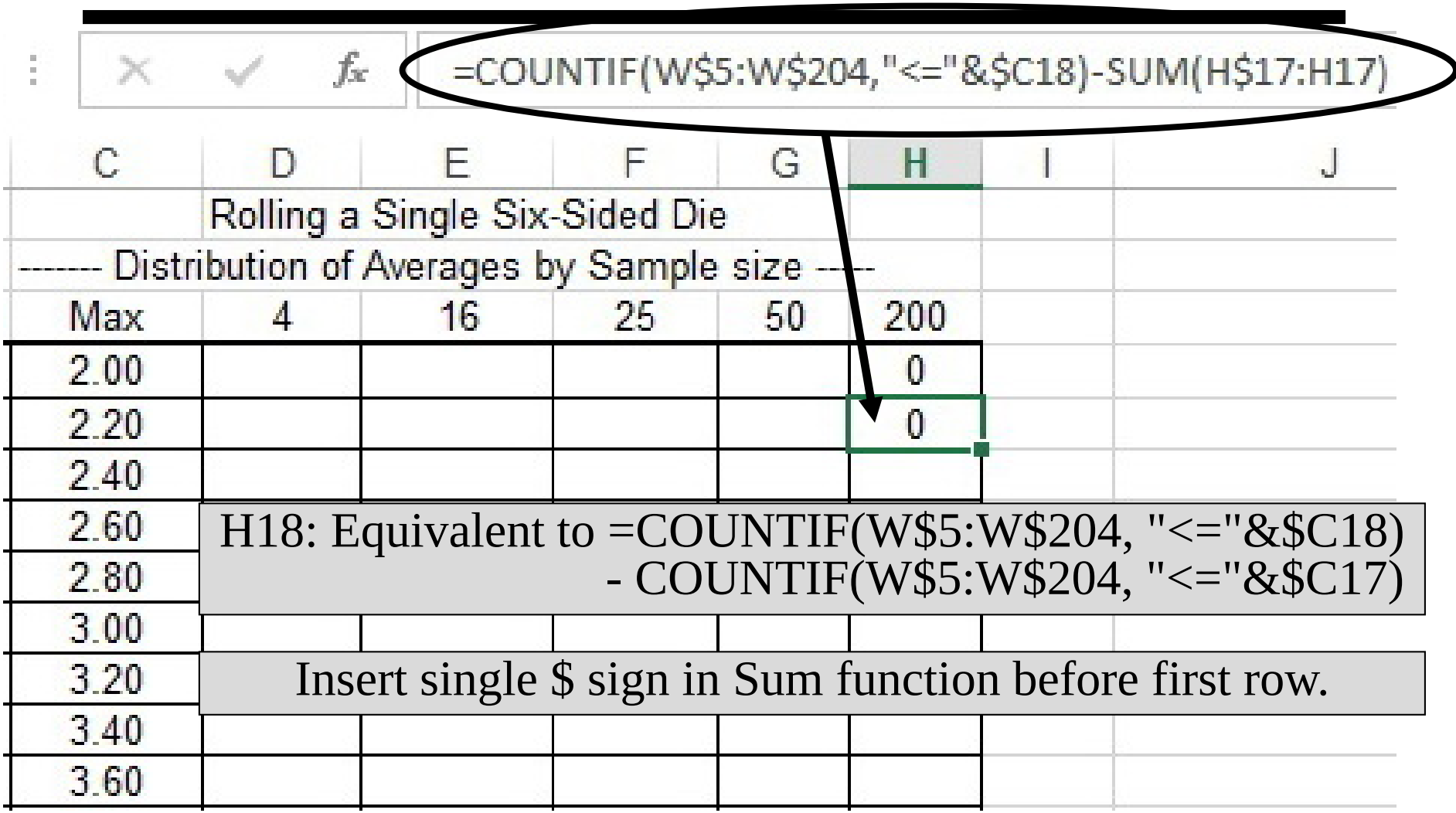
	A	B	C	D	E	F	G	H
1	A	B	C	D	E	F	G	H
2	③		Population: Six sides of a Single Die					
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Population			0.748	StdDev / Half-R	
6								
7	④	Sample Size		4	16	25	50	200
8		Mean of Means		3.52	3.51	3.52	3.53	3.52
9		Median of Means		3.50	3.53	3.52	3.52	3.51
10		Std Dev of Means		0.85	0.41	0.34	0.24	0.12
11		SE: % of Pop SD		50%	24%	20%	14%	7%
12		1/Sqrt(n)		50%	25%	20%	14%	7%

5a: Insert COUNTIF function in H17



Insert $ sign in before Column in range; before Row in single cell!

5b: Insert "=CountIF() –Sum()" function in 2nd row: H18



5c: Drag H18, =CountIF() –Sum(), formula down to bottom row

H18 =COUNTIF(W$5:W$204, "<="&$C18)-SUM(H$17:H17)

	A	B	C	D	E	F	G	H	I	J
14	⑤			Rolling a Single Six-Sided Die						
15			------- Distribution of Averages by Sample size -----							
16		Midpoint	Max	4	16	25	50	200		
17		2.10	2.00					0		
18		2.10	2.20					0		
19		2.30	2.40					0		
20		2.50	2.60					0		
21		2.70	2.80					0		
22		2.90	3.00					0		
23		3.10	3.20					3		
24		3.30	3.40					44		
25		3.50	3.60					117		
26		3.70	3.80					36		
27		3.90	4.00					0		
28		4.10	4.20					0		
29		4.30	4.40					0		
30		4.50	4.60					0		
31		4.70	4.80					0		
32		4.90	5.00					0		
33		Total # of Samples		0	0	0	0	200		

Drag entire right column (H17:H32) left-ward to fill out frequency table.

Insert SUM at bottom of each column.

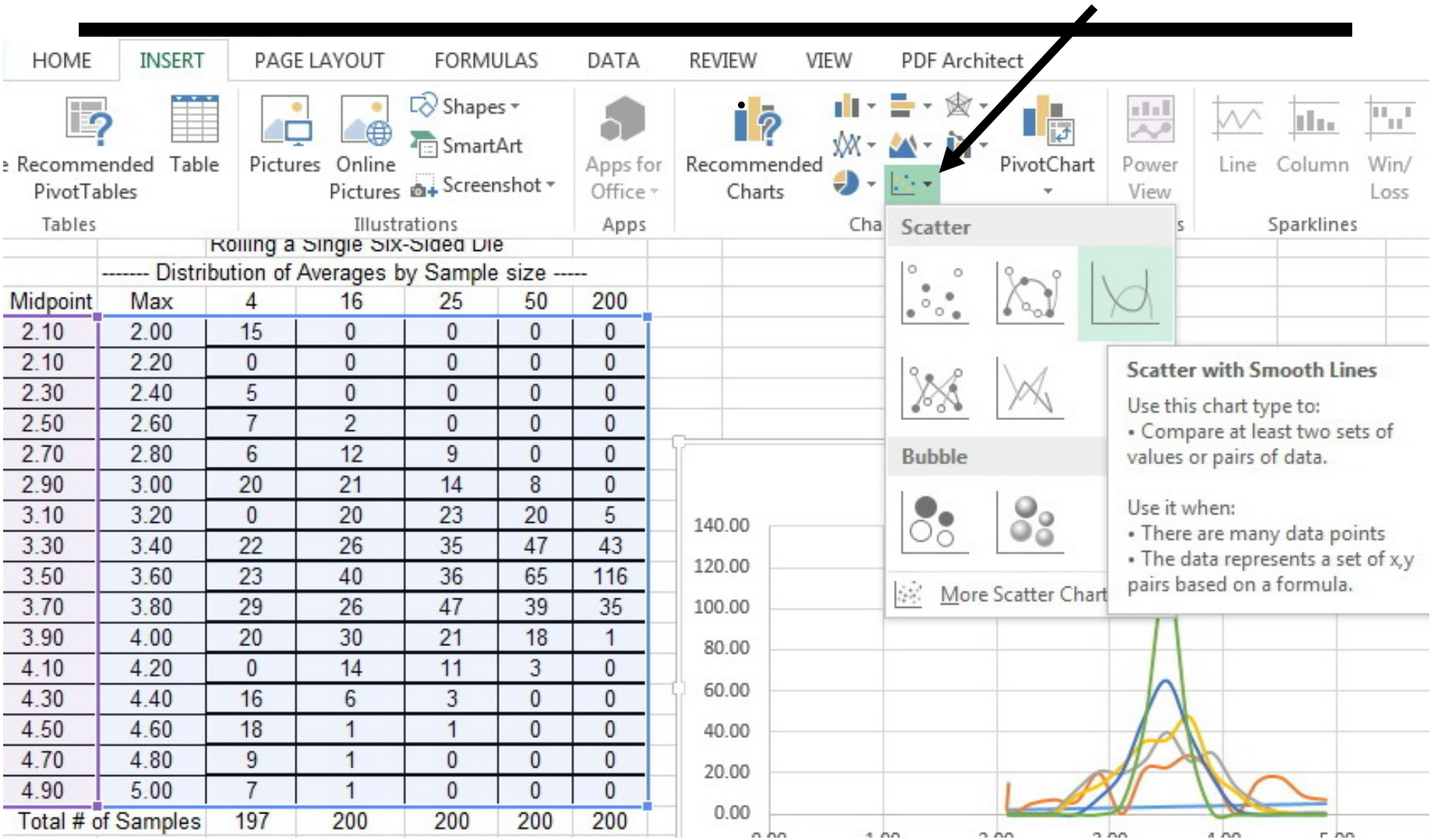
May be different from 200.

Do not include H17 when dragging H18 downward!!!

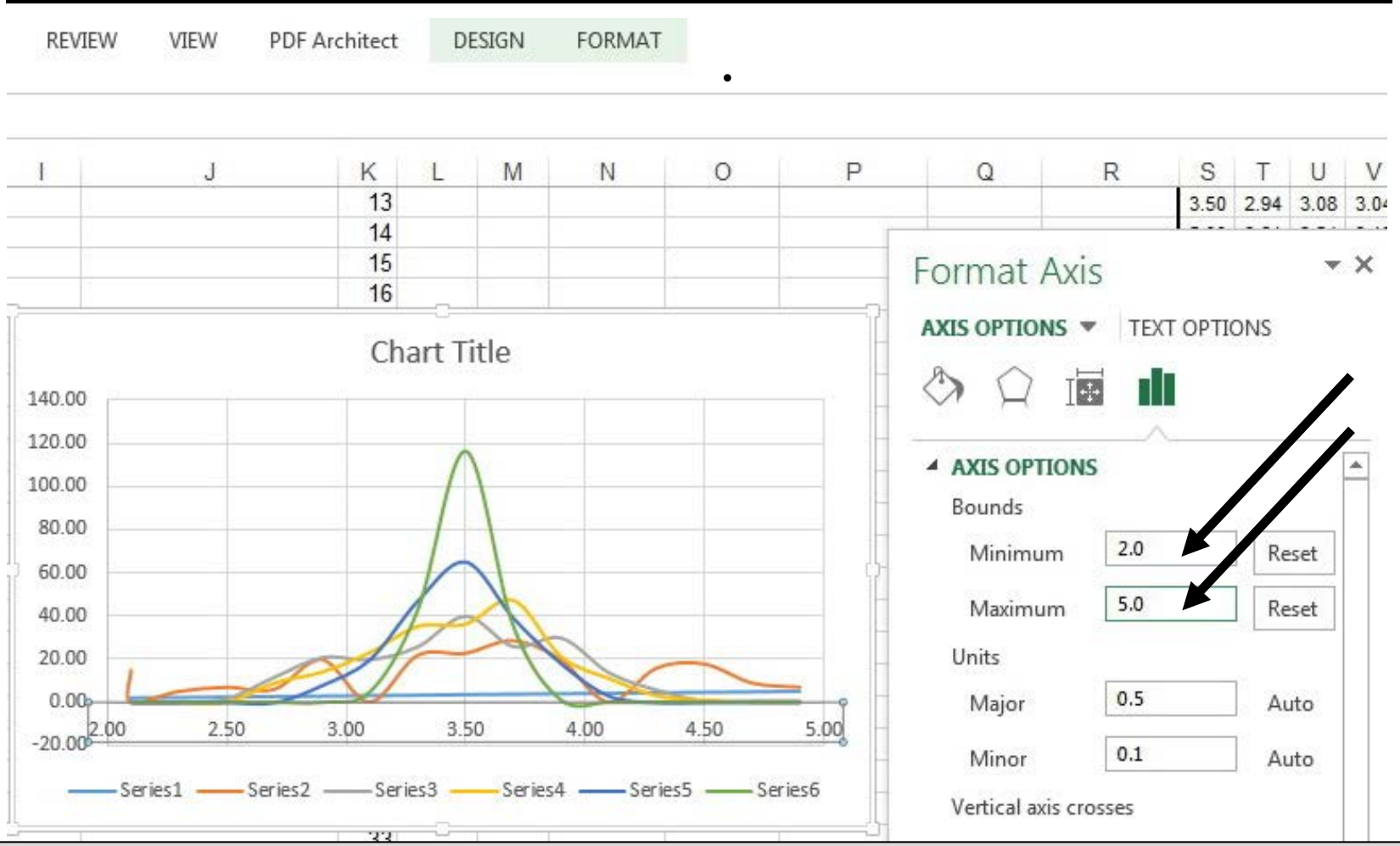
6a: Select Data (B17:H32) to use in Histogram

	A	B	C	D	E	F	G	H
14	⑤			Rolling a Single Six-Sided Die				
15				------- Distribution of Averages by Sample size -----				
16		Midpoint	Max	4	16	25	50	200
17		2.10	2.00	10	0	0	0	0
18		2.10	2.20	0	0	0	0	0
19		2.30	2.40	10	0	0	0	0
20		2.50	2.60	8	3	0	0	0
21		2.70	2.80	8	4	3	0	0
22		2.90	3.00	24	18	20	2	0
23		3.10	3.20	0	23	18	20	1
24		3.30	3.40	34	34	40	50	39
25		3.50	3.60	17	33	53	66	113
26		3.70	3.80	21	33	33	46	47
27		3.90	4.00	21	30	18	15	0
28		4.10	4.20	0	14	13	1	0
29		4.30	4.40	13	8	2	0	0
30		4.50	4.60	15	0	0	0	0
31		4.70	4.80	11	0	0	0	0
32		4.90	5.00	4	0	0	0	0
33		Total # of Samples		196	200	200	200	200

6b: Insert X-Y Scatter Plot



6c: Format Horizontal Axis: Change Min and Max



Delete legend.

6d: Select/delete Series 5 (size 50), 3 (size 16) and 1 (Max).

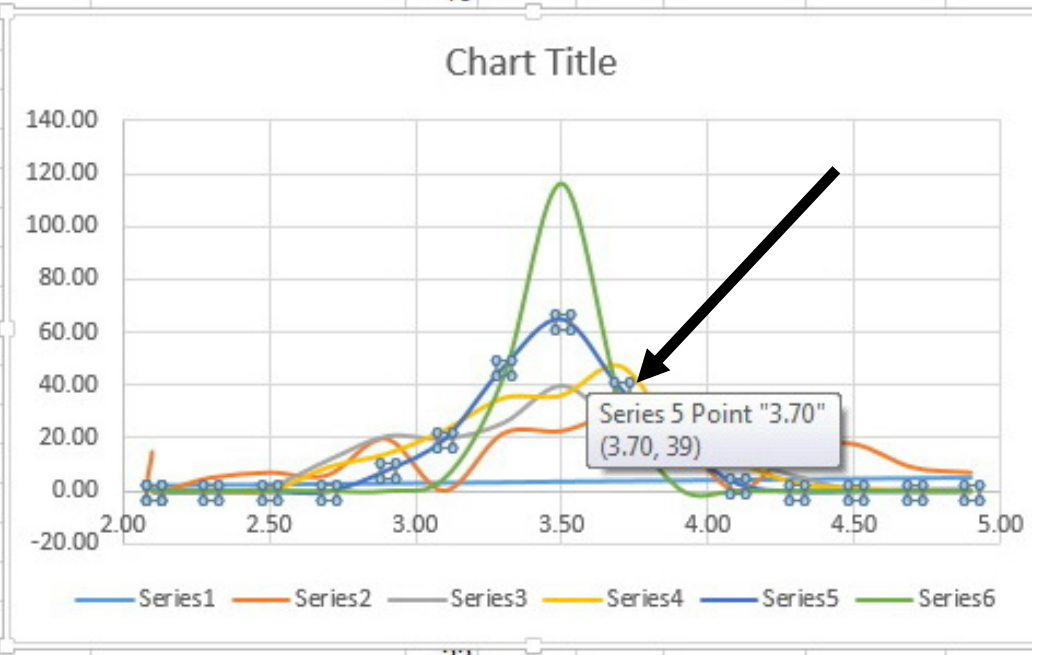
HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW PDF Architect DESIGN FORMAT

=SERIES(,Data!$B$17:$B$32,Data!$G$17:$G$32,5)

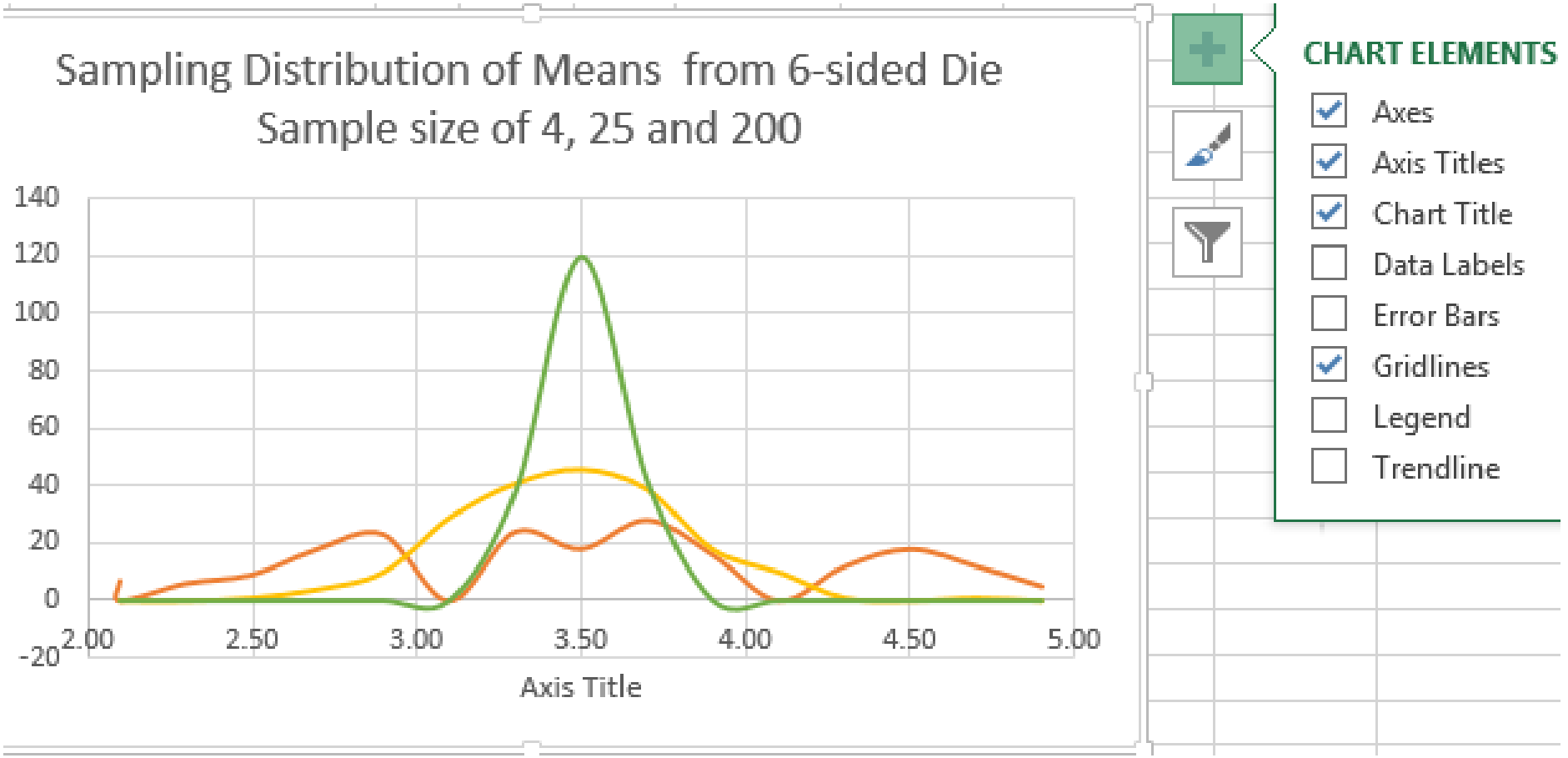
Rolling a Single Six-Sided Die

------- Distribution of Averages by Sample size -----

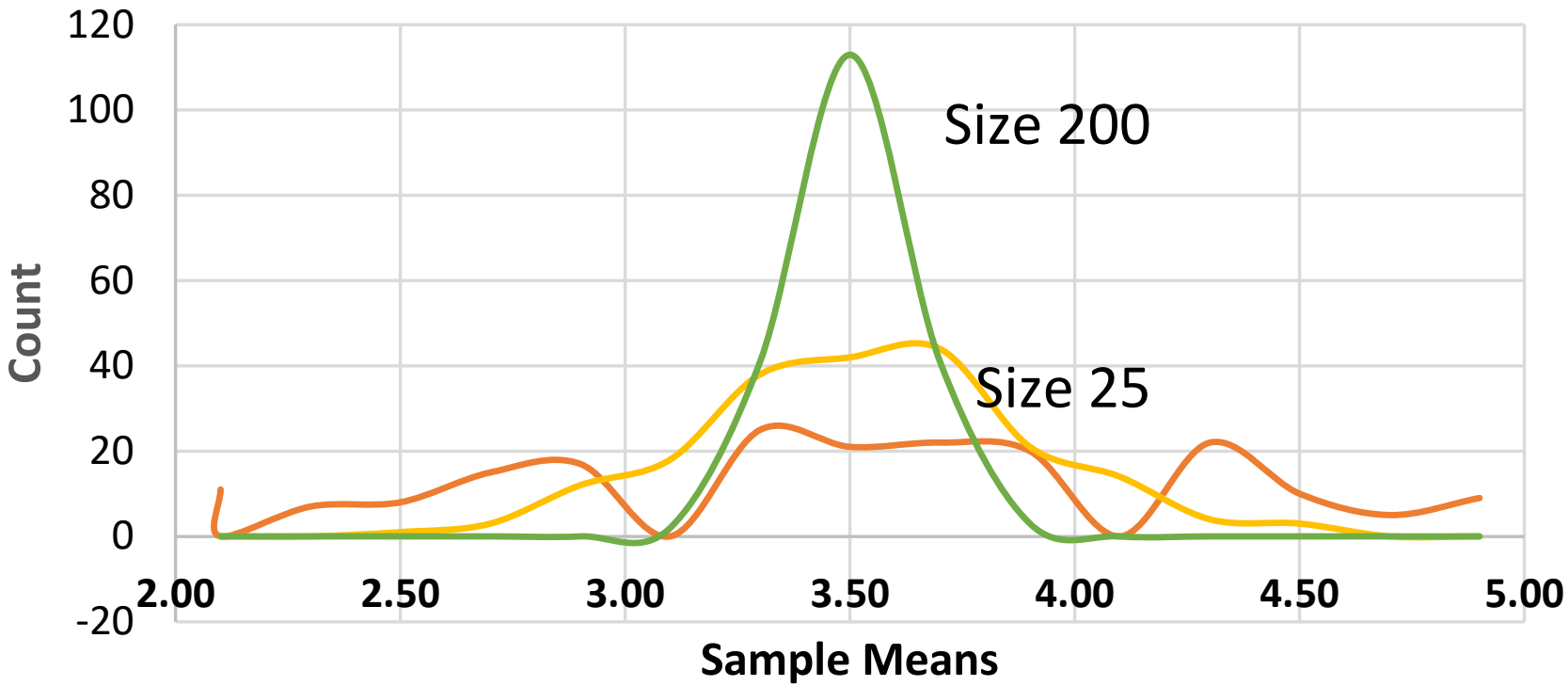
Midpoint	Max	4	16	25	50	200
2.10	2.00	15	0	0	0	0
2.10	2.20	0	0	0	0	0
2.30	2.40	5	0	0	0	0
2.50	2.60	7	2	0	0	0
2.70	2.80	6	12	9	0	0
2.90	3.00	20	21	14	8	0
3.10	3.20	0	20	23	20	5
3.30	3.40	22	26	35	47	43
3.50	3.60	23	40	36	65	116
3.70	3.80	29	26	47	39	35
3.90	4.00	20	30	21	18	1
4.10	4.20	0	14	11	3	0
4.30	4.40	16	6	3	0	0
4.50	4.60	18	1	1	0	0
4.70	4.80	9	1	0	0	0
4.90	5.00	7	1	0	0	0
Total # of Samples		197	200	200	200	200



6e: Insert Title and Horizontal Axis Text



6f: Insert legends in text boxes. Final Result



Summary

When sampling from a process, the population "size" is "infinite". That doesn't influence the standard deviation.

Notice as sample size increases, the standard error (the std. deviation of the sample means) quickly decreases – as a percentage of the population standard deviation.

A sample of size 4 is expected to have a standard error that is only a half of the population standard deviation: a sample of 25 has a fifth, a sample of 100 has a tenth and a sample of 10,000 has a hundredth.