

VOG 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

VOG 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

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

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

VOG 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.

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

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

Randomly generated numbers will be random!

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

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

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

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

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

`=AVERAGE($AA5:AP5)`

L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	aa ab
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																

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
Row Averages						
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

`=AVERAGE(B3:G3)`

A	B	C	D	E	F	G
A	B	C	D	E	F	G
Population: Six sides of a Single Die						
1	1	2	3	4	5	6
4	3.5	Mean				3.5
5	1.71	Std Dev Population				StdDev / Half-Range

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

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

`=STDEV.P(B3:G3)`

A	B	C	D	E	F	G	H	I
A	B	C	D	E	F	G	H	I
Population: Six sides of a Single Die								
1	1	2	3	4	5	6		
4	3.5	Mean				3.5	Median	
5	1.71	Std Dev Population					StdDev / Half-Range	
7	4	Sample Size	16	25	50	200		
8		Mean of Means						
9		Median of Means						

In cell F5, insert `=B5/((G3 - B3)/2)`

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

`=AVERAGE(W$5:W$204)`

B	C	D	E	F	G	H
B	C	D	E	F	G	H
Population: Six sides of a Single Die						
1	1	2	3	4	5	6
3.5	Mean				3.5	Median
1.71	Std Dev Population			0.683	StdDev / Half-R	
	Sample Size	4	16	25	50	200
	Mean of Means					3.50
	Median of Means					3.50
	Std Dev of Means					0.12
	SE: % of Pop SD					
	1/Sqrt(n)					

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

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

`=STDEV.S(W5:W204)`

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

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

H11: $=H10/SB5$

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 $=1/\text{SQRT}(H7)$ into cell H12

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

D11: $=D10/SB5$

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

H17: $=\text{COUNTIF}(W\$5:W\$204, "<="&\$C17)$

A	B	C	D	E	F	G	H	I
13								
14	3	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!

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

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

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 $=\text{COUNTIF}(W\$5:W\$204, "<="&\$C18) - \text{COUNTIF}(W\$5:W\$204, "<="&\$C17)$

Insert single \$ sign in Sum function before first row.

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

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

A	B	C	D	E	F	G	H	I	J
14	3	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	3	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

19

6b: Insert X-Y Scatter Plot

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	0	0
3.10	3.20	0	20	23	20	5
3.30	3.40	22	26	35	47	43
3.50	3.60	23	49	38	65	118
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

20

6c: Format Horizontal Axis: Change Min and Max

Chart Title

Delete legend.

21

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

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	0	0
3.10	3.20	0	20	23	20	5
3.30	3.40	22	26	35	47	43
3.50	3.60	23	49	38	65	118
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

22

6e: Insert Title and Horizontal Axis Text

Sampling Distribution of Means from 6-sided Die
Sample size of 4, 25 and 200

Axis Title

23

6f: Insert legends in text boxes. Final Result

Sampling Distribution of Sample Means: 6-side Die
Sample sizes of 4, 25 and 200

Count

Sample Means

Size 200

Size 25

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.