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
- Excel2015-Sampling-TDIe-Demo-Sindes.pdf
- A picture of the output is at www.StatLit.org/pdf/ Excel2013-Sampling-1Die-Demo-Output.pdf

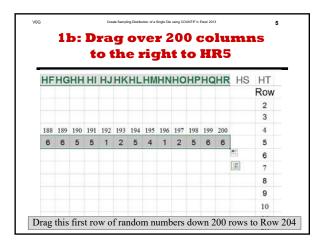
Steps in Creating Sampling Distributions for a single die

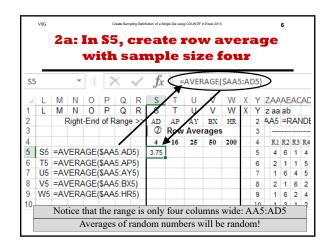
tribution of a Single Die using COUNTIF in Excel 2013

- 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.

Create line-graph histograms by sample size.
Upload completed worksheet.

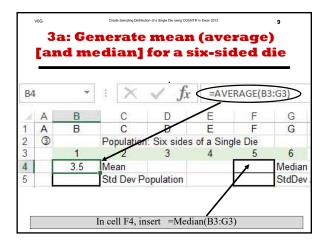
V0G	;	٥	reate Sampling Distri	bution of a Sin;	yle Die using (COUNTIF	in Excel 20	113			4	
	1a:	Inse	ert R unc						n(1	۱,6)	
-					C							
AA5		*	X	\checkmark	Ĵx		=R	AND	BET	WEEI	N(1,	5)
	U	V	W	Х	Υ	Ζ	AA	AB	AC	AD	ĄĖ	AF
1	U	V	W	Х	Υ	z	aa	ab			/	
2					2							
3	Aver	ages			3				+			
4	25	50	200		4		R1	R7	R3	R4	R5	R6
5					5		1	K				
6	R	andomly	genera	ted nu	imbei	rs w	/ill b	e ran	dom	!		

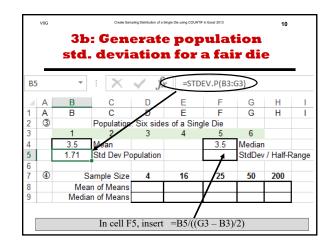


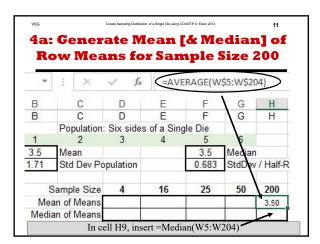


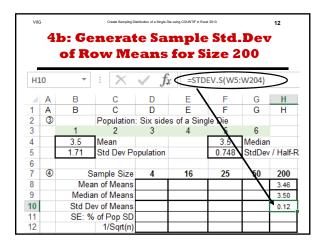
		2 b							te						ıg	e			
T5			*					Ťx	le						>				
	L	М	N	0	P	Q	R	S	\int_{T}	U		W	Х	Y	ZAA	AE	AC	AD	AE
1	L	М	N	0	Ρ	Q	R	8	Т	U.	/v	W	Х		z aa				
2		F	light-	End	l of F	Rang	e >>,	AD	(AP)	AY	BX	HR		2	4 A5	=R	AN	DE	BE
3			Ŭ					2	Row	Ave	rages			3					
4							/	4	16	25	50	200		4	R1	R2	R3	R4	R5
5	S 5	=AV	ERA	GE	\$AA	5:AI	D5)	3.25	3.69					5	3	3	2	5	4
6	T5	=AV	ERA	GE	SA4	5.AI	P5)							6	6	3	3	2	1
7	U5	=AV	ERA	GE	\$AA	5:A	Y5)							7	2	2	2	1	4
8	V5	=AV	ERA	GE	SA4	(5:B)	(5)							8	4	4	2	5	5
9	W5	=AV	ERA	GE	\$AA	5:HI	R5)							9	1	5	2	5	4
10														10	5	2	2	1	4
11		ss F9												11	6	6	5	2	1
	In	U5:	W5,	cre	ate	row	ave	rage	s for	sam	ple s	izes	25	, 50) and	12	$0\overline{0}$)	

G		ate Sampling Dis						8
2 a :	: Drag	F 12	sti	tov	A	/e	rag	es,
S	55:W5,	, do	wn	to	Ro	W	204	-
	S	Т	П	V	W	X	Y	
	S	T	Ŭ	V	W	X	Y	
	AD	AP	AY	BX	HR	1	2	
	2	Row	Ave	rages			3	
	4	16	25	50	200		4	
	3.25	3.56	3.16	3.26	3.40		5	
	12						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	







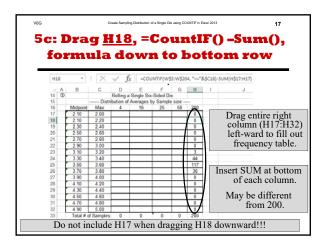


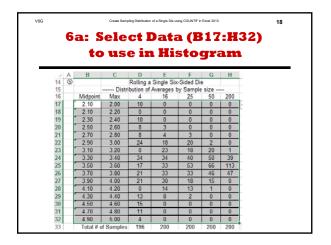
								13
	_		11 R a					
	:0	гор	ulati	on 5	ta. 1	Jevia	atio	n
H	1	-	$: \times$	√ fs)/\$B5		
	A	в	С	D	Е	F	G	Н
1	Α	В	С	D	E	F	G	Н
2	3		Population	Six side	s of a Sin	gle Die		
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Po	pulation		0.748	StdDev	/ Half-F
6							1	
7	4	S	ample Size	4	16	25	50	200
8		Mea	an of Means					3.47
9		Media	an of Means					3.50
10		Std D	ev of Means					0.12
11		SE: %	6 of Pop SD					7%
12			1/Sqrt(n)					*
_		т	sert $=1/S$	ODT/II	7)	cell H12		

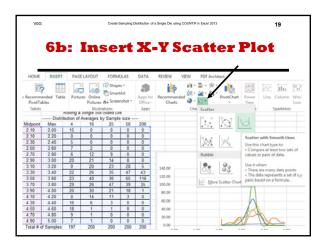
	V0G		Create Sample	ng Distribution of a Sin	gle Die using COUNTIF in E	Excel 2013		14
		4d :	Drag	H8:	H12 c	olur	nn	
		left-v	vard f	to fil	lout	the t	abl	•
D1	11	*	$: \times$	\sqrt{f}	. =D10/	(\$B5		
	А	в	С	D	F	F	G	н
1	A	B	C	D	E	F	G	H
2	3		Population	: Six side	s of a Sing	le Die		
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Po	pulation		0.748	StdDev	/ Half-R
6								
7	4	S	ample Size	4	16	25	50	200
8		Mea	n of Means	3.52	3.51	3.52	3.53	3.52
9		Media	n of Means	3.50	3.53	3.52	3.52	3.51
10		Std De	v 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%

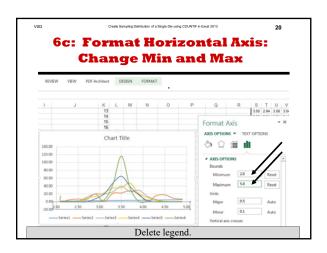
VO	G		Create Sar	npling Distribution of a	a Single Die using COUN	TIF in Excel 2013			15							
	5:	- In	sert	COI	INTI	F fu	nc	tion	•							
									-							
				in	H17											
H1	.7	*	: 🗙	√ f.	x =cou	UNTIF(W\$	\$5:W\$20)4, "<="8	\$C17)							
	А	В	С	D	Е	F	G	Н	I							
13	~															
14	6	Rolling a Single Six-Sided Die														
15						· ·		<u> </u>								
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		1		1									
In	sert	\$ sign in	before (Column	in range;	before I	Row in	single	cell!							

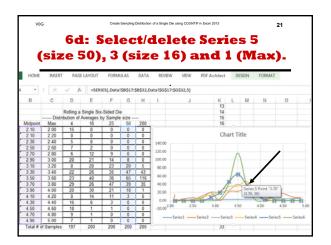
V0G		Create Sampli	ng Distribution of a Sin	gle Die using COU	JNTIF in Excel 201:	3	16
5 b						-Sum : H18	0"
\times	$\checkmark f_x$	=COL	JNTIF(W\$	\$5:W\$20)4,"<="8	\$C18)-SUN	1(H\$17:H17)
С	D	E	F	G	Н	I	J
	Rolling a	Single Six	-Sided Di	e			
Dist	ribution of /				-		
Max	4	16	25	50	200		
2.00	· · · · · · · · · · · · · · · · · · ·				0		
2.20					0		
2.40							
2.60	H18: Ec	uivalent	to $=CO$	UNTIF	(W\$5:	W\$204. "<	<="&\$C18)
2.80		1	- CO	UNTIF	(W\$5:	W\$204, "<	(="&\$C17)
3.00		-	1	1			, ,
3.20	Inse	ert single	\$ sign in	1 Sum	function	n before fi	rst row.
3.40	T	0		1	1		
3.60	Y 27	2			9 S		

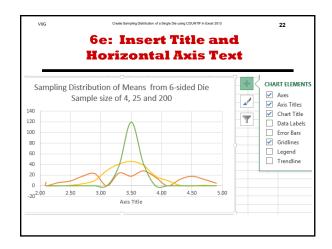


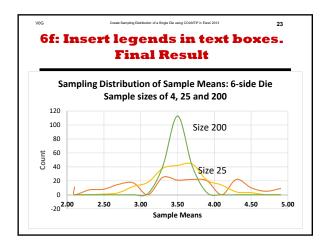


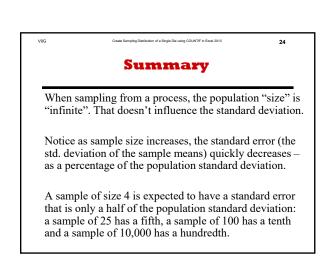












Create Sampling Distributions from a Single Die in Excel 2013

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

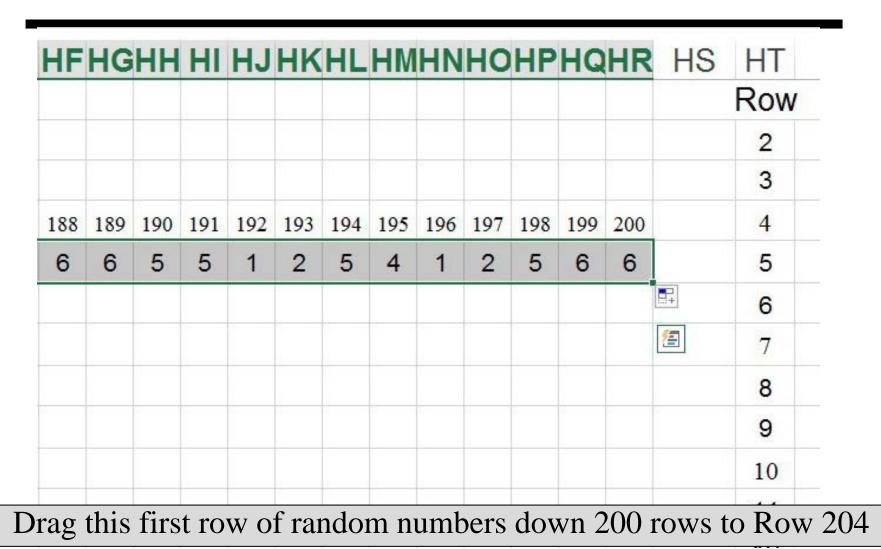
Access Pre-formatted Data Worksheet:

- 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.

AA5		* :	×	1	fx		=R/	AND	BET	NEEI	N(1.6	5)
				-	J.,							<u> </u>
	U	V	W	Х	Υ	Ζ	AA	AB	AC	AD	ĄÉ	AF,
1	U	V	W	Х	Υ	z	aa	ab				
2					2							
3	Aver	ages			3				+			
4	25	50	200		4		R1	R2	R3	R4	R5	R6
5					5		1	K				
6					0							
	R	andomly	generat	ed nu	ımber	'S W	vill b	e ran	dom	!		

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



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

S 5			Ŧ	:	>	<	\checkmark	f _x		AVER	AGE(\$AA5	5:AI	D5)	>			
	L	М	N	0	Р	Q	R	S	T	U	X	W	Х	Y	ZAA	AE	AC	AD
1	L	Μ	N	0	Р	Q	R	ß	Т	Ų	V	W	Х	Υ	z aa	ab		
2		F	Right	-End	of F	lang	e >>	AD	AP	AY	BX	HR		2	4 A5	=R	۱A	ID E
3								0	Row	Aver	rages			3				
4								4	16	25	50	200		4	R1	R2	R3	R4
5	S5	=A\	/ER/	\GE(\$AA	5:AE	05)	3.75						5	4	6	1	4
6	T5	=A\	/ER/	AGE(\$AA	5:AF	P5)							6	2	1	1	5
7	U5	=A\	/ER/	\GE(\$AA	5:A)	(5)							7	1	6	4	5
8	V5	=A\	/ER/	\GE(\$AA	5:B)	(5)							8	2	1	6	2
9	W5	=A\	/ER/	\GE(\$AA	5:HF	R 5)							9	1	6	2	4
10 ₁			•		-			-	0	-				10	4	3	4	2
		No	tice	that	the	rang	ge is	only	tour	colu	mns	wide	: A	A5	O:AL)5		
			A	Aver	ages	s of	rand	om r	umb	ers w	vill b	e ran	do	m!				

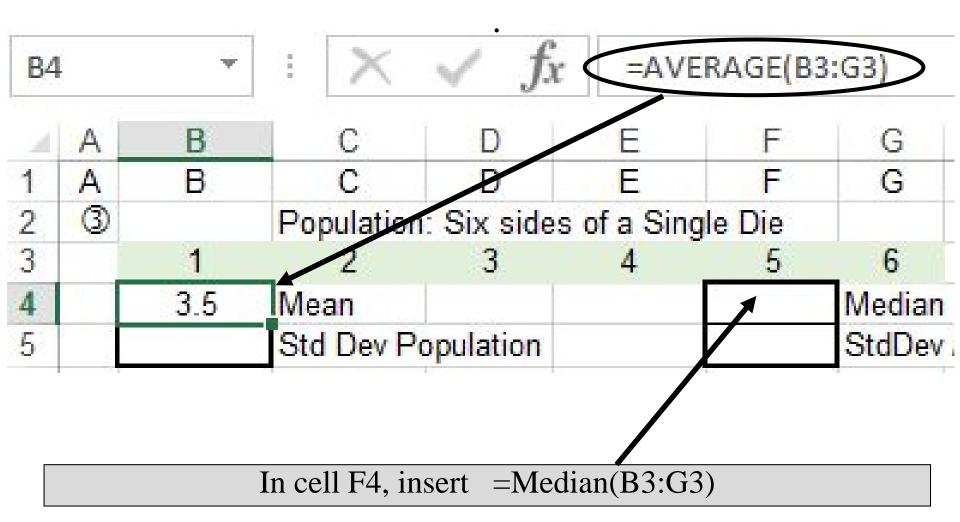
2b: In T5, create row average with sample size of 16

Т5			Ŧ		2	×	\checkmark	f x		AVER	RAGE(\$AA5	5:Al	P5)	>				
	L	М	Ν	0	Р	Q	R	S	Т	U		W	Х	Y	ZAA	AE	AC	AD	AE,
1	L	Μ	N	0	Ρ	Q	R	ß	Τ	U	V	W	Х	Υ	z aa	ab			
2		F	Right	-End	l of F	Rang	e >>	AD	(AP)	AY	BX	HR		2	4 A5	=F	٩A	IDE	3ET
3								2	Row	Aver	rages			3					
4								4	16 🖌	25	50	200		4	R1	R2	R3	R4	R 5
5	S 5	=A\	/ER/	AGE(\$AA	.5: AE)5)	3.25	3.69					5	3	3	2	5	4
6	T5	=A\	/ER/	AGE(\$AA	5:AF	⁹ 5)							6	6	3	3	2	1
7	U5	=A\	/ER/	AGE(\$AA	.5:Α <mark>`</mark>	(5)							7	2	2	2	1	4
8	V 5	=A\	/ER/	AGE(\$AA	.5:ΒX	(5)							8	4	4	2	5	5
9	W5	=A\	/ER/	AGE(\$AA	.5:HF	₹5)							9	1	5	2	5	4
10														10	5	2	2	1	4
11	Pres	ss F	9 to g	get n	ew c	lata.								11	6	6	5	2	1
	In	U5:	W5	, cre	ate	row	ave	erage	s for	sam	ple s	izes	25	, 50) an	d 2	200)	
				Rig	ght-	end	of r	ange	s are	show	wn ir	n S2:	W	2					

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

S	T	U	V	W	Х	Y
S	Т	U	٧	W	Х	Y
AD	AP	AY	BX	HR		2
0	Row	Ave	rages	1		3
4	16	25	50	200	4	×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



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

B5		-	$: \times$	V ß	=STD	EV.P(B3:	G3		
	A	В	С	D	E	F	G	Н	I
1	Α	В	С		Е	F	G	Н	
2	3		Population	Six side	s of a Sing	gle Die			
3		1	2	3	4	5	6		
4		3.5	Mean			3.5	Median		
5		1.71	Std Dev Po	pulation		4	StdDev	/ Half-F	Range
6									
7	4	S	ample Size	4	16	25	50	200	
8		Mea	n of Means						
9		Media	n of Means			Y			

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

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

-	: ×	$\checkmark f_x$	=AV	ERAGE(W	\$5:W\$2	04)				
В	С	D	E	F	G	Η				
В	С	D	E	F	G	Н				
Population: Six sides of a Single Die										
1	2	3	4	5	6					
3.5	Mean			3.5	Median					
1.71	Std Dev Po	pulation		0.683	StdDev	/ Half-R				
ç	Sample Size	4	16	25	50	200				
Me	an of Means	S				3.50				
Medi	an of Means									
	In ce	ell H9, ins	ert =Medi	an(W5:W	204)					

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

H1	.0	Ŧ	: 🗙	$\checkmark f$	f_x =STDEV.S(W5:W204)								
	А	В	С	D	E	F	G	H					
1	Α	В	С	D	E	F	G	Н					
2	3		Population	: Six side	s of a Sing	le Nie							
3		1	2	3	4	Ţ	6						
4		3.5	Mean			3.5	Median						
5		1.71	Std Dev Po	opulation		0.748	StdDev	/ Half-R					
6													
7	4	Sa	ample Size	4	16	25	50	200					
8		Mear	n of Means					3.46					
9		Media	n of Means					3.50					
10		Std De	v of Means					0.12					
11		SE: %	of Pop SD					Ī					
12			1/Sqrt(n)										

12

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

H1	1	Ŧ	: X	$\checkmark f$	r =H10,	(\$B5		
	Α	В	С	D	E	F	G	Н
1	Α	В	С	D	E	F	G	Н
2	3		Population	: Six side	s of a Sing	le Die		
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Po	opulation		0.748	StdDev	/ Half-R
6								
7	4	S	ample Size	4	16	25	50	200
8		Mea	n of Means					3.47
9		Media	n of Means					3.50
10		Std De	v of Means					0.12
11		SE: %	of Pop SD					7%
12			1/Sqrt(n)					→

Insert =1/SQRT(H7) into cell H12

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

D1)11 -		: ×	$\checkmark f$. =D10/	/\$B5		
	А	В	С	D	Е	F	G	Н
1	Α	В	С	D	E	F	G	Н
2	3		Population	: Six side	s of a Sing	le Die		
3		1	2	3	4	5	6	
4		3.5	Mean			3.5	Median	
5		1.71	Std Dev Po	opulation		0.748	StdDev	/ Half-R
6								
7	4	Sa	ample Size	4	16	25	50	200
8		Mear	n of Means	3.52	3.51	3.52	3.53	3.52
9		Media	n of Means	3.50	3.53	3.52	3.52	3.51
10		Std De	v 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

H1	7	Ŧ	: 🗙	$< \sqrt{f_x} = \text{COUNTIF}(W$5:W$204, "<="8")$										
4	Α	В	С	D	E	F	G	Η	I					
13														
14	6			Rolling a Single Six-Sided Die										
15			Distri	Distribution of Averages by Sample size										
16		Midpoint	Max	4	16	25	50	200						
17		2.10	2.00					0	8					
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											
In	sert	t \$ sign in	before C	Column i	in range;	before F	Row in	single	cell!					

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

\times	√ f ₃	a =col	UNTIF(W\$	5:W\$20)4,"<="8	&\$C18)-9	SUM(H\$17:H17)
С	D	E	F	G	Η		J
	Rolling a	Single Six	е				
Dist	ribution of	Averages I	by Sample	e size			
Max	4	16	25	50	200		
2.00			_	_	0		
2.20					0		
2.40							
2.60	H18: E	quivalent	$t = CO^{2}$	UNTIF	F(W\$5:	W\$204	, "<="&\$C18)
2.80		1	- CO	UNTIF	F(W\$5:	W\$204	, "<="&\$C17)
3.00							
3.20	Inse	ert single	\$ sign in	Sum	functio	n befor	e first row.
3.40							
3.60							

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

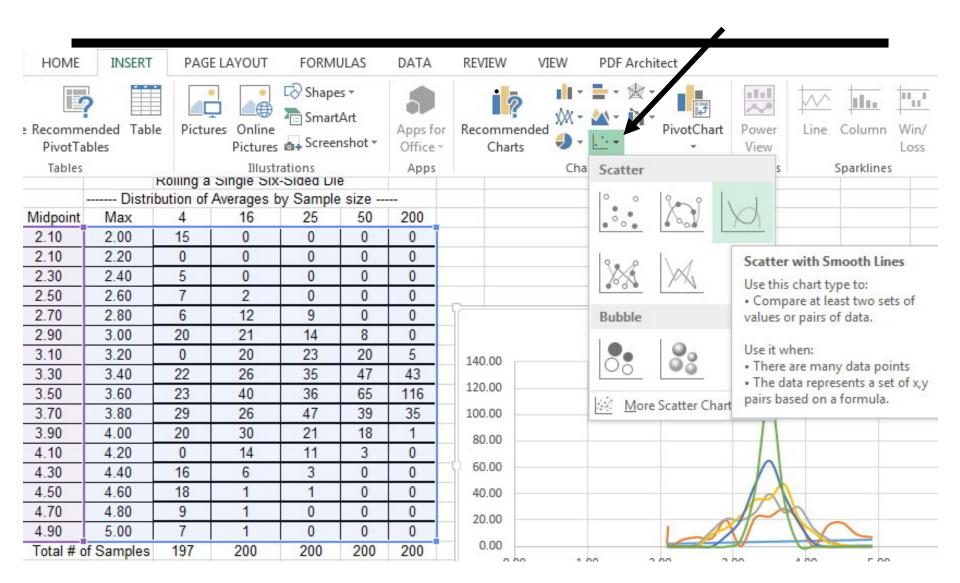
H1	8	*	: 🗙	✓ f.	τ =CO	UNTIF(W\$	5:W\$20)4, "<="8	\$C18)-SUM(H\$17:H17)
	Α	В	С	D	E	F	G	H	1	J
14	5			Rolling a	Single Size	x-Sided Di	е			
15			Distri	bution of	Averages	by Sample	e size			
16		Midpoint	Max	4	16	25	50	200		
17	_ [2.10	2.00							Drag entire right
18		2.10	2.20					0		
19	1	2.30	2.40					0		column (H17:H32)
20		2.50	2.60					0		left-ward to fill out
21		2.70	2.80					0		
22		2.90	3.00					0		frequency table.
23		3.10	3.20					3		1 v
24		3.30	3.40					44		
25		3.50	3.60					117		
26		3.70	3.80					36		Insert SUM at bottom
27	1	3.90	4.00					0		
28	1	4.10	4.20					0		of each column.
29	1	4.30	4.40					0		
30		4.50	4.60					0		May be different
31		4.70	4.80					0		
32		4.90	5.00							from 200.
33		Total # o	f Samples	0	0	0	0	200		

Do not include H17 when dragging H18 downward!!!

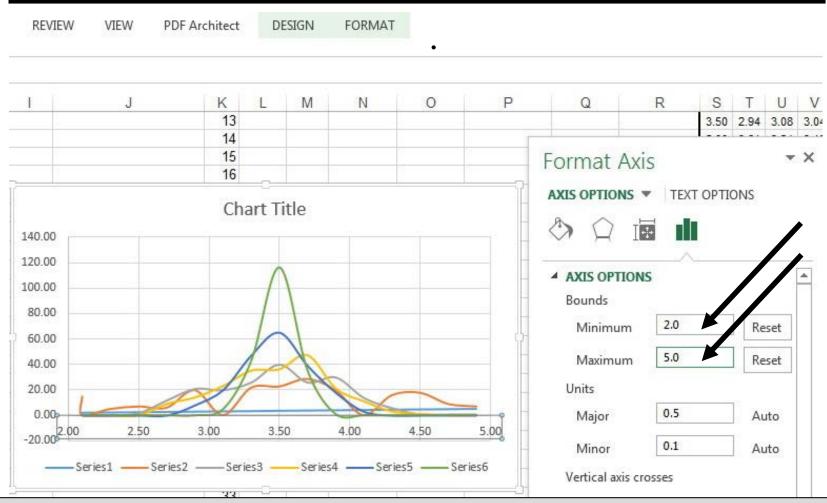
6a: Select Data (B17:H32) to use in Histogram

	А	В	С	D	E	F	G	Н
14	6			Rolling a	Single Six-			
15			Distri	bution of	Averages b	y 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 # o	of Samples	196	200	200	200	200

6b: Insert X-Y Scatter Plot



6c: Format Horizontal Axis: Change Min and Max



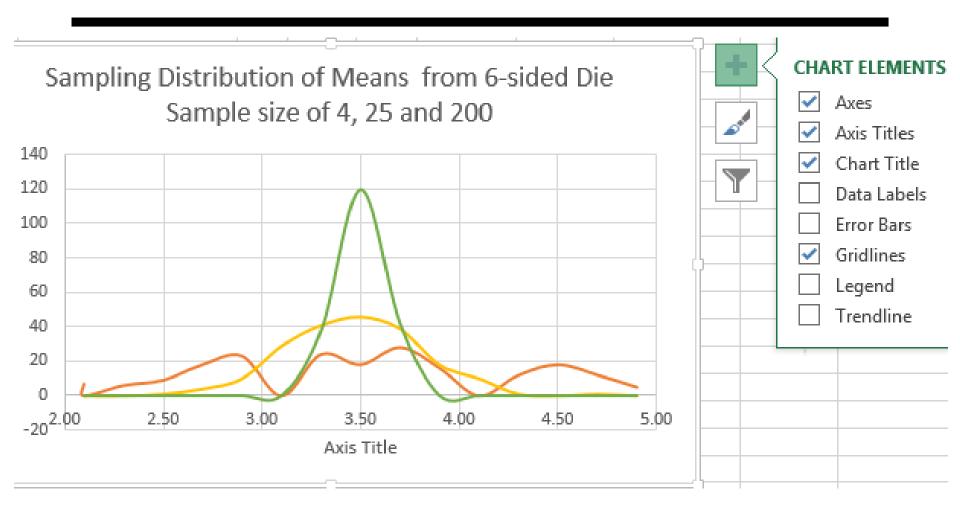
Delete legend.

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

HOME	INSERT	PAGE	LAYOUT	FORM	ULAS	DATA	REVIEW	VIEW	PDF Archite	ect	DESI	GN	FORMAT		
4 *	: ×	$\checkmark f_x$	=SER	IES(,Data	!\$B\$17:	\$B\$32,D	ata!\$G\$17:\$0	3\$32,5)							
В	С	D	E	F	G	Н	1	J	K		L	М	N	0	<u> </u> - 10
										13					
-			Single Six			<u></u>				14					
	Distri									15					
Midpoint	· · · · · · · · · · · · · · · · · · ·	4	16	25	50	200	C			16					
2.10	2.00	15	0	0	0	0				cL -		2			
2.10	2.20	0	0	0	0	0	-			Cha	rt Titl	е			
2.30	2.40	5	0	0	0	0	140.00				-		1		
2.50	2.60	7	2	0	0	0	120.00								
2.70	2.80	6	12	9	0	0					\wedge				
2.90	3.00	20	21	14	8	0	100.00				/ \				
3.10	3.20	0	20	23	20	5	80.00				1				
3.30	3.40	22	26	35	47	43					1 22				
3.50	3.60	23	40	36	65	116	60.00			0.	100				
3.70	3.80	29	26	47	39	35	40.00			A	0	10	•	-	
3.90	4.00	20	30	21	18	1	20.00			4	4		5 Point "3.70"		
4.10	4.20	0	14	11	3	0	1			%	l	(3.70,			
4.30	4.40	16	6	3	0	0	0.00		0-0 0-0		2.50		38 88	0-0 0-0 0-0 0-0 4 E 0	0-0
4.50	4.60	18	1	1	0	0	-20.002.00	2.50	0 3.00		3.50		4.00	4.50	5.00
4.70	4.80	9	1	0	0	0			Series2	Series	-3	Series	A Series	5 0	eries 6
4.90	5.00	7	1	0	0	0		11031	JG 1032	3010	55	30103	54	5 5	CI ICSU
Total #	of Samples	197	200	200	200	200				33					

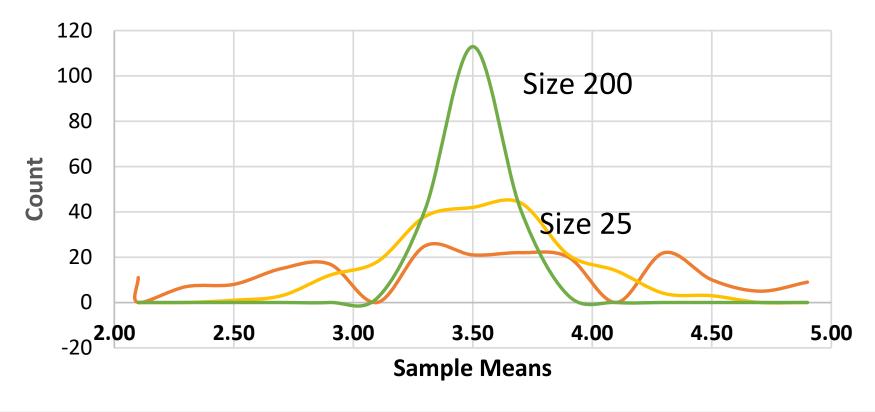
22

6e: Insert Title and Horizontal Axis Text



6f: Insert legends in text boxes. Final Result

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





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.